DYNAMICS OF LABOR MARKET EARNINGS AND SECTOR OF EMPLOYMENT IN URBAN MEXICO, 1987-2002.

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Robert Duval Hernández

January 2006

DYNAMICS OF LABOR MARKET EARNINGS AND SECTOR OF EMPLOYMENT IN URBAN MEXICO, 1987-2002.

Robert Duval Hernández, Ph.D.
Cornell University 2006

This dissertation studies labor earnings mobility in the short-run and the structure of labor markets in urban Mexico, from 1987 to 2002.

In the first part it gauges the average earnings mobility in the economy and whether mobility equalized longer-term earnings. It also analyzes whether the mobility patterns differ by groups of the population, and whether mobility reduced longer-term earnings inequality between and within groups. The groups considered are age, education, gender, quintile of initial earnings, sector, and region groups.

In general, average earnings mobility fluctuated around zero, with the exception of the late eighties and early 2000, when individuals experienced gains, and of the years following the 1994 Peso crisis, when individuals experienced large losses. These patterns are shared by the majority of the groups of the population, with the exception of initial earnings quintile and sector groups. For these groups, the most advantaged individuals experience the largest losses, while the most disadvantaged ones experience the largest gains. Furthermore, mobility equalized longer-term earnings for the entire population during most of the periods studied, and it helped reduce longer-term earnings inequality within-groups. However, mobility only sometimes equalized longer-term earnings between groups.

The second part of the dissertation studies short-run earnings dynamics at the

individual level. In particular, it examines whether mobility benefits more the initially advantaged individuals. Regression analysis shows a high level of convergence between the earnings of rich and poor. However, part of this convergence reflects transitory adjustments in earnings. In practice, most of the individuals keep their permanent advantage, leading to little convergence between rich and poor. The major exception to this finding occurs in the aftermath of the Peso crisis, when individuals with a high permanent advantage experienced greater losses than the rest of the population. The *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility is gauged. Education, gender, region and transitions between sectors are important factors affecting earnings mobility.

The final part of the dissertation tests whether Mexican urban labor markets are segmented between formal and informal sectors. An econometric structural model of sector choice is estimated, and a strong evidence of rationing of formal sector jobs is found. The estimations also show that individuals rationed out of the formal sector would experience large gains by moving into the formal sector.

BIOGRAPHICAL SKETCH

I was a born in Mexico City on May 4th 1977. Son of a Mexican historian and a Haitian doctor/epistemologist, I was raised with a great love and respect for academics, but also with a constant guilt about the injustices that occurred in the developing world.

My decision to study Economics was grounded on two factors. First, Economics was the discipline that provided the perfect mix of social sciences and mathematics. Second, having been raised in the Marxist belief that the "structure" (i.e. the economy) determined the "super-structure" (i.e. the institutions, culture, etc.), I thought that, if ever I wanted to make a difference from an academic position, Economics was the best place to be.

At the end I changed my beliefs, and now I think that institutions and norms are equally important, if not more, than the economic structure. Also, I stopped believing that my academic work can actually make any difference in fixing the life of the poor in the world. Anyway, I still think Economics is a great discipline, with interesting questions and methods.

I obtained my undergraduate degree in Economics in CIDE in 1999, and the following year I came to Cornell University to do my Ph.D.

I married (3 times) Nelia Charalambous in 2004, and in 2005 I was blessed with two nephews, Renato and Camilo.

A mi familia Guy, Dolores, Karina y Dolores,

y a Nelia con toda mi alma.

ACKNOWLEDGEMENTS

Even when this dissertation is officially authored by one single person, behind it there is the support of many, without whom this work would have not come to a completion.

I was extremely privileged in having a great Dissertation Committee.

My thesis advisor, Gary S. Fields, provided continuous support during my Ph.D. studies. It was him who insisted that I do a dissertation on mobility, especially at moments when I wanted to abandon the whole topic. He was always willing to spend time talking to me, and displayed the patience of a saint when discussing academic issues. I learned a lot from him, both about economics and about life in general. In addition to this, he and his wife Vivian made their house a cozy home for us students.

I was also extremely lucky in having the support of Kaushik Basu and George Jakubson. They are great economists who care deeply about their profession. Talking to them taught me many things, and inspired me in many ways. Their teachings both in the classroom and outside of it, have shaped my way of thinking Economics, and go beyond their contribution to this dissertation.

Chapter 4 benefited from the comments of seminar participants at Cornell University, Universidad Autónoma de Guanajuato, CIDE, and Universidad Autónoma de Nuevo León.

Funding from CONACYT and Cornell University is gratefully acknowledged.

I want to thank my parents in law Antigoni and Sofocli Charalambous for cheering up hard for me finishing, and for taking care of me during the last days of production of this dissertation.

Thanks go to my friends Mabel Andalón, Pablo Aznavurián, Vladimir Barberena, Juan Carlos Chávez, Fernando Coda, Jean Cortissoz, Emilio "Chacal" Domínguez, Ryuichi Fukuoka, Perla Ibarlucea, Rodrigo Mendes, Monica Ponce, Andrea Pozas, Julio Antonio Ríos, Fernando Schwartz, Hugo Solís, Gabriel Torres, Juan Carlos Trejo, Hector Valdés, Víctor Valdés, and Omar Zamora, for their warm heart and the good times we spent together.

My family Guy, Dolores, Karina, and Dolores, always gave me love and support in my endeavors. Although 200 pages of Labor Economics do not make up for 5 years of absence, they know that even at a distance my thoughts and love have always been with them. Thanks also go to Hugo and Manuel for their friendship and kindness, and to Renato and Camilo for bringing such joy to the family.

Finally, I would like to thank my partner Nelia Charalambous. Por su amor, paciencia y cuidados, pero sobre todo por la belleza de su compañía.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 Motivation

The topic of this dissertation is the study of the short-run dynamics of labor market earnings and sector of employment in urban Mexico from 1987 to 2002.

The research on economic mobility issues in developing countries is fairly recent. While mobility studies were performed in the developed world since the second-half of the twentieth century, such topics started to be addressed in developing nations only towards the end of last century. Part of this neglect was due to empirical limitations, and part of it due to conceptual limitations. Empirically, very few longitudinal studies (that follow the same unit of analysis over time) had been conducted on these economies, and since mobility studies depend crucially on this type of data, there were few opportunities to make such research. However, the conceptual limitations of a discipline too used to analyzing issues of inequality and poverty from a cross-sectional perspective also played a role in this story. Even when development economists were concerned with issues of mobility, their habit of relying on cross-sectional empirical data, prevented them from tackling many of the interesting mobility questions once new data became available.[1]

The importance of mobility studies comes from their ability to follow the destinies of the agents of interest (be it individuals, households, firms, etc.) over time. This advantage over cross-sectional studies helps in tackling new questions that are inherently dynamic in nature. For instance, questions like "How did the

[1] As evidence to this, comes the fact that the data used in this dissertation started becoming available in the late eighties. However, it was not until very recently, that it was used to conduct a proper mobility study in Mexico.

initially poor fare over time?", "What employment transitions did an individual experience?", "Has long-run inequality increased or decreased?", etc. can only be answered with panel data.

The present dissertation devotes two of its main chapters to the study of earnings mobility *per se*. The other main chapter deals with the old controversial issue of whether labor markets are segmented, and analyzes the implications of this market structure on earnings mobility.

The data used in this dissertation is a series of short overlapping panels with quarterly information tracking individuals for at most 1 year. The geographic area studied is urban Mexico, and the period covered by these short-lived panels goes from January 1987 to December 2002. The dataset used is a household survey that provides information on labor market variables (including labor market earnings) and the socioeconomic characteristics of individuals.

Although it would be desirable to have panel data following individuals for more than a year, such data does not exist yet for Mexico. Instead, what can be learned from this study is the short-run dynamics of earnings and sector of employment, under varying macroeconomic conditions. Having many short-lived panels, covering over 15 years, is a unique opportunity for analyzing an economy like the Mexican, which underwent radical transformations during this period, including a long process of economic liberalization, and a severe financial crisis in December 1994.

1.2 Questions

The questions addressed in this dissertation can be organized in three main categories.

1) Aggregate Earnings Mobility Questions:

"What are the average earnings gains and losses in the economy?", "Are these earnings mobility patterns the same for different groups of the population?", "Does mobility equalize earnings over time?", "Does mobility equalize earnings *within* groups over time?" and "Does mobility equalize earnings *between* groups over time?". The analysis is performed on various groups of the population according to age, education, gender, initial earnings quintile, region and sector groups.

2) Initial Earnings and the Determinants of Earnings Mobility:

"How does initial earnings affect earnings mobility?", in particular "Are the most advantaged individuals gaining more (losing less) in terms of earnings changes?", "What is the *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility?", and "How do these factors affect the impact of initial earnings on mobility?".

3) Labor Market Segmentation and Mobility:

"Are Mexican labor markets segmented?", in particular "Are formal sector jobs rationed?". If so, "What are the potential earnings gains that rationed individuals could experience by moving into the formal sector?", "What is the *actual* earnings mobility experienced by rationed individuals if they manage to enter the formal sector in further periods?".

All these questions will be addressed using the same survey. The earnings mobility analyzed will be yearly earnings mobility. This allows capturing the longest period of time possible for each individual, and avoids having to worry about issues of seasonality. Also, these questions will be analyzed under varying macroeconomic conditions, capturing the various changes experienced by the Mexican economy during the years under study.

1.3 Previous Research and My Contribution

The study of economic mobility is relatively new in economics. It is mostly during the second half of the twentieth century that this area of research sprung in the empirical literature. Part of the reason for this late entrance into the discipline was the lack of suitable longitudinal data following individuals over time as well as the evolution of their incomes, consumption patterns, employment status, occupation, etc. Although earnings mobility studies started relatively early within the mobility literature, with the seminal 'permanent earnings model' of Friedman and Kuznets (1954), it would take two more decades for such studies to start appearing more systematically.[2]

In the case of developing economies the study of mobility is even more recent, with most of the studies dating from the last decade and a half. Again, the lack of suitable panel data is partly to blame for this. For many decades the attention of empirical economists interested in issues of welfare, poverty and inequality, was solely focused on static pictures of the economy. Needless to say, this was rather limiting, since it prevented answering questions like whether some individuals were stuck in poverty, or what caused upward mobility.

Mobility studies have been recently performed for a wide variety of developing countries like China (Nee, 1996; Jalan and Ravallion, 2000), and India (Gaiha, 1988; Coondoo and Dutta, 1990) in Asia, South Africa (Fields *et al.*, 2003a,b), and Zimbabwe (Gunning *et al.*, 2000) in Africa, Hungary (Galasi, 1998) in Eastern Europe, and Argentina (Sánchez-Puerta, 2005), Peru (Herrera, 1999), and Venezuela (Freije, 2001) in Latin America, just to mention a few examples. This expansion

[2] Atkinson *et al.* (1992), and Fields and Ok (1999a) present an overview on the methods and empirical results relevant for developed economies.

can also be witnessed by the fact that two major journals specializing in Development Economics have devoted entire issues to the topic (August 2000 issue of the *Journal of Development Studies*, and March 2003 issue of *World Development*).[3]

For the case of Mexico, the number of mobility studies addressing earnings dynamics is limited. Previous studies focusing on the evolution of incomes (or earnings) were made by means of comparable cross-sections and not with longitudinal data (see for instance Lustig and Székely (1999), and Cortés (2000)). Many of these studies focused on the evolution of poverty and inequality over time, but since they did not follow the same individuals they do not constitute mobility studies as such. A large body of literature built around studying the impact of the (quite recurrent) past economic crises, and the strategies used by households to cope with them. Examples of this literature are Selby and Browning (1992), López-Acevedo and Salinas (2000), Attanasio and Székely (2004), McKenzie (2003), Maloney and Cunningham (2000) and Maloney *et al.* (2004). However, the vast majority of these references are not mobility studies in a strict sense.[4]

It is important to mention that there is a set of studies on economic mobility in Mexico that analyzes other types of variables like education (Binder and Woodruff (2002), Dahan and Gaviria (1999), Behrman *et al.* (2001)), occupation (Latapí, 1992), sector of economic activity[5] (Ibarlucea, 2003), and regional convergence in earnings (Aguayo-Tellez, 2005).

Aggregate earnings mobility in Mexico has been studied in only a couple of papers. The studies by Wodon (2001) and Yitzhaki and Wodon (2002) examine

[3]Further references are reviewed in Fields (2001) and Baulch and Hoddinott (2000).

[4]The exception being the papers coauthored by Maloney, which will be reviewed in detail in the literature review section of Chapter 4.

[5]Meaning manufacturing, agriculture, services, etc.

Time dependence in economic positions, the first comparing urban Mexico and Argentina, the second only in rural Mexico.[6] None of these papers provide evidence on aggregate Directional mobility and Mobility as an equalizer of longer-term earnings. Hence, the results that be presented in this dissertation relative to these two mobility concepts, are new to the literature.

On the study of mobility at the individual level, two papers recently appeared focusing on conditional and unconditional convergence in earnings over time (see Antman and Mckenzie, 2005a,b). These are studies for urban Mexico and use the same data as this dissertation. However, the way they use the data to analyze mobility is quite different from the route taken here. More precisely, the authors opted for not using the panel structure of the data, and instead of tracking individuals over a year, they constructed pseudo-panels by following age-education cohort groups. The earnings mobility they analyze is therefore between-cohort mobility. Although their approach has some advantages in getting them a longer coverage of time (a given cohort can be followed for as many years as there are surveys), and reducing the potential negative effects of measurement error and attrition bias, their method is not without problems. In particular, by tracking the mobility of cohorts they fail to analyze any intra-cohort mobility that might take place. Second, one cannot be sure that the mobility experienced by a cohort group actually represents the true mobility experienced by a given group of individuals. Problems like migration, deaths, and household dissolution and creation, can lead to incorrect inferences when this method is applied.

In the light of these studies, the chapter which studies the role of initial earnings and the determinants of earnings mobility contributes to research in the field by

[6] There is an ongoing research study on mobility and political outcomes in Mexico by Fernanda Arce and Luis Felipe Lopez Calva.

focusing on the short-run earnings dynamics experienced by individuals in Mexican urban areas. It also provides further evidence on the role played by socioeconomic factors in determining mobility, and interprets the results obtained within the framework of a structural model of earnings. Finally, that chapter tests the robustness of the findings to different measures of initial advantage, to measurement error, and to attrition bias.

Regarding the topic of labor market segmentation, there is a vast literature studying this issue for Mexico and the world. This literature is reviewed in section 5.2 and the details will not be discussed here. However, it must be said that much of this literature was based on descriptive statistics that are suggestive at best, or when a formal structural analysis was performed to test for segmentation, many of the econometric models did not explicitly separate the decision of the worker applying for a formal sector job, from the one of the employer hiring the applicant.

The contribution of Chapter 5 is to use the segmentation test proposed in Pisani and Pagán (2003) to the Mexican case, and to extend their application by estimating selectivity-corrected earnings equations. This segmentation test has the advantage of explicitly modeling the decisions of workers and employers. Also, the estimation of selectivity-corrected earnings equations helps to generate counterfactual earnings for the individuals rationed out of the formal sector. These predicted formal sector earnings provide a measure of the potential earnings gains that these individuals would experience if they were to move to the formal sector. In addition to that, this chapter exploits the panel structure of the data to provide evidence on the sectoral transitions and the *actual* earnings mobility experienced by individuals who were restricted from entering the formal sector in the initial period.

1.4 Structure of the Dissertation

The structure of the dissertation is as follows. In Chapter 2 an overall presentation of the dataset used in this dissertation is made. Chapter 3 contains the analysis of the aggregate mobility concepts of Directional mobility and Mobility as an equalizer of longer-term earnings. This is the first substantive chapter in the dissertation. Chapter 4 contains the model and results analyzing the impact of initial earnings on mobility and the determinants of earnings mobility. Finally, Chapter 5 contains the tests of segmentation in labor markets and the evidence of its impact on sectoral and earnings mobility. Chapter 6 concludes.

Chapter 2

Data

2.1 Introduction

This dissertation uses data coming from the National Survey of Urban Employment (in Spanish "Encuesta Nacional de Empleo Urbano") from now on abbreviated as ENEU. This is a survey conducted on Mexican urban households with the purpose of inquiring about the conditions that prevail in urban labor markets. This chapter introduces the characteristics of the ENEU, and presents some descriptive statistics for the sample under study. This will provide the reader a panorama of the evolution of earnings, as well as of the general macroeconomic conditions in Mexico during the period under study.

2.2 Description of the Survey

As previously mentioned, the ENEU is a survey conducted on Mexican households with the purpose of inquiring about the conditions that prevail in urban labor markets. The database is a rotating panel with quarterly data. It tracks individuals for at most 5 quarters in the most important urban areas of the country. The sampling is done in three stages, based on a sampling frame of dwellings.

The survey gathers information about socioeconomic characteristics such as age, gender, education, marital status, labor force participation, labor market earnings, sector of employment, occupation, type of fringe benefits, hours worked in the market, as well as hours devoted to other activities (e.g., housework), type of employment contract, firm size, employment search activity, dwelling characteristics, etc. The survey is designed to be geographically and socioeconomically representative of urban Mexico.[1] Furthermore, it is one of the surveys used by the government to create employment statistics.

Although the geographic coverage of the ENEU has expanded substantially over time, the analysis performed in this dissertation restricts the sample to the 16 cities that originally appeared in the sample of 1987. Doing otherwise might confound the “true” evolution of mobility measures with the effects caused by the expansion of the geographical coverage.

As previously mentioned, this is a study on earnings mobility in the short-run. The short temporal coverage of each panel makes it impossible to draw conclusions on what happens with earnings mobility in the long-run. However, a long period of time, that runs from 1987 to 2002, is covered by using many of these short-lived overlapping panels. These years include several periods of growth and the major recession following the 1994 Peso crisis. This period also coincides with the years of trade liberalization in Mexico.[2]

In the 3rd quarter of 1994 a new questionnaire was applied in the survey, and although there were minor changes with respect to the previous questionnaire, this dissertation avoids using the panels in which the individuals questioned experienced a change in questionnaire.

The unit of analysis in the following chapters is an individual worker. Throughout the panel individuals are matched according to their household and personal

[1]By the latter it is meant that the survey stratifies the population according to wealth. Hence representative results can be obtained for each of these strata.

[2]Unfortunately, there are no comparable datasets that cover the periods previous to the beginning of the trade liberalization process. Also, since this process was a slow and continuous one, it is hard to find a break point which would identify the impact of the new trade conditions, separately from other factors.

identification numbers, and also by their age, gender and years of education in order to minimize the probability of spurious matching.

This dissertation focuses most of its attention to studying one-year earnings mobility (from the initial interview quarter to the same quarter next year). Since the survey follows individuals for at most 5 quarters, then at most one observation of yearly earnings changes exists per individual. This precludes using any panel data econometric technique. Although there is information available on earnings at other quarters, studying earnings changes over shorter periods of time (e.g., quarterly mobility) is not pursued here. The reason for not doing this, is that a year is already a short period over which to study mobility. Further extensions to the present work can include modeling the covariance structure of earnings using all the 5 periods available for each individual.

The subpopulation of study is restricted to individuals between 25 and 60 years of age. Also, with the exception of Chapter 5, which starts with some cross-sectional estimates, all the estimations are restricted to individuals who are double-labor-force participants (i.e., that are in the labor force both in the first interview and one year later). The reason for applying these restrictions is to avoid having to analyze the mobility associated with first time entries into the labor force by young people who recently graduated from school, retirement decisions, and transitions-in-and-out of the labor force in general. Although this might hide some interesting effects, like the entrance into the labor force of family members during times of recessions, it helps focusing the dissertation on the earnings mobility experienced by workers who are more permanently attached to the labor market. Note however, that unemployed individuals *are* included in the analysis. This is done because finding and/or losing a job is an important event *per se*, that affects the welfare of an individual and it involves a significant change in earnings.

The earnings variable is real earnings measured in 2002 Mexican pesos. The advantage for using such year as base period, is that back then the nominal exchange rate between the US Dollar and the Mexican peso was about 10 pesos per dollar, something which facilitates the interpretation of the results to the international reader.

Finally, it is important to remark that all the calculations here presented are weighted estimations using the survey factor weights. This is done in order to obtain estimates that are more representative at the national urban level. The author has performed most of these estimations with unweighted data, and the conclusions reached do not change signinificantly.[3] In the unweighted estimations, the directions of the main parameters of interest remain unchanged, although the magnitudes of the numbers reported sometimes differ.[4] Also, whenever possible the standard error estimates and statistics calculated are adjusted for the characteristics of the survey design, in particular for clustering and stratification.

The next section presents some descriptive statistics for the main sample under study in the ENEU, and some data on the macroeconomic evolution of the Mexican economy. All the calculations presented were performed by the author using the ENEU surveys, unless explicitly noted.[5]

[3]Some of these results are included in Fields *et al.* (2005). These results will not be included here in order to avoid doubling the length of this dissertation.

[4]Due to problems of attrition in the panel sample the weighted mobility estimators are not necessarily representative. Further extensions of this research will perform the same estimations using post-stratified weights to see if there is any variation in the conclusions reached.

[5]The descriptive statistics here presented pertain to the sample used in the estimations of chapters 3 and 4. Chapter 5 uses a slightly different sample, for which the descriptive statistics are included in the appendix to that chapter.

2.3 Descriptive Statistics

The first graph introduced in this section presents the evolution of GDP in Mexico together with the evolution of average earnings for the sample that will be used for most of the dissertation. Figure 2.1 shows an upward trend in real GDP during the years going from 1987 to 1994, when the December Peso crisis hit the economy. After this crisis, output suffered a sharp downturn, out of which it started rapidly recovering. Nevertheless, the pre-1994 aggregate output levels were not recovered until 1999. From 1999 onwards, the Mexican economy continued its growth, but by 2001-2002 a new recession had started again.

Average earnings followed the steady growth of the economy during the period going from 1987 to 1994. After the 1994 Peso crisis, earnings fell dramatically. However, unlike GDP, they did not start their recovery but until much later. It was only after 1999 that average earnings started growing again, and by 2002 they hadn't reached yet their pre-1994 level. These graphs serve to illustrate that earnings did not exactly match the evolution of aggregate output. Instead, three clear periods can be distinguished in the evolution of earnings. The first period goes from the 1st quarter of 1987 to the 2nd quarter 1993, the second going from the 3rd quarter 1994 to the 1st quarter 1999, and the last one going from the 2nd quarter 1999 to the 4th quarter of 2002. This classification of the evolution of earnings will become useful later for presenting results in a more compact way, by pooling panels for these broad periods.

The evolution of earnings inequality in the sample is presented in Figure 2.2, which depicts the Gini coefficient estimated at the initial interview of the panels. This picture shows that inequality grew for the first half of the period under study. After 1994 it remained fairly constant, though at this higher level.

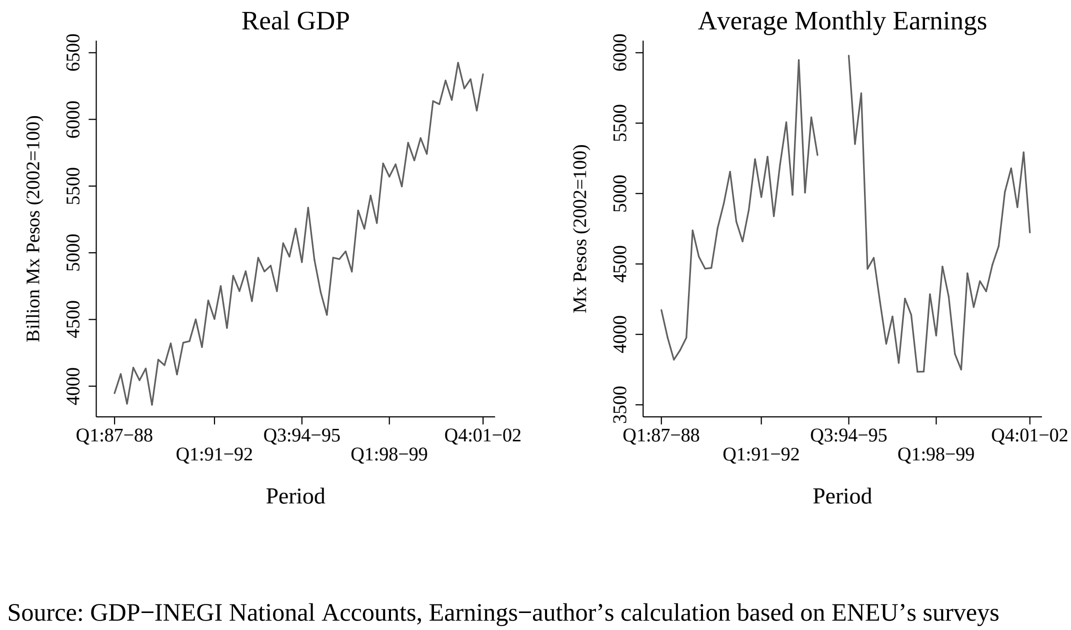


Figure 2.1: GDP and Average Earnings

Figure 2.2: Gini Coefficient for Individuals in the Sample

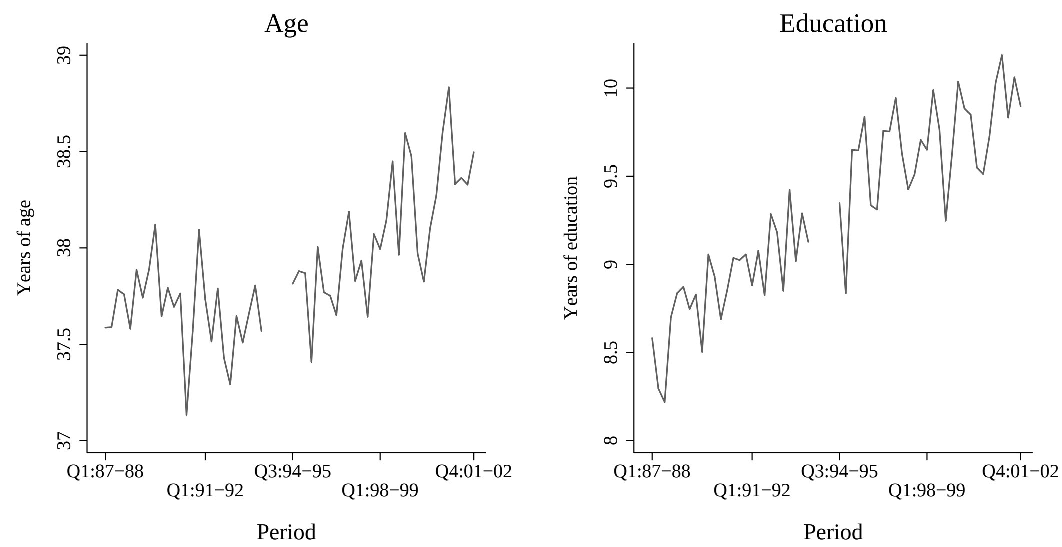


Figure 2.3: Average Age and Education

The evolution of other characteristics of the sample are plotted in Figures 2.3-2.5. Figure 2.3 contains information on the average age and education for the individuals in the sample. It is clear from these pictures that age remained fairly constant around 37 years of age for the first half of the period, but then increased afterwards, raising the average age in the sample by 1 year. On the other hand, education steadily increased from 8 1/2 to 10 years of education by the end of the sample.

It is important to recall that these numbers pertain to individuals from 25 to 60 years of age who are in the labor force both at the first and the last interview. Hence, these numbers do not represent the whole labor force.[6]

[6]Including all the individuals in the labor force would have the effect of reducing the average age and bringing in less educated workers.

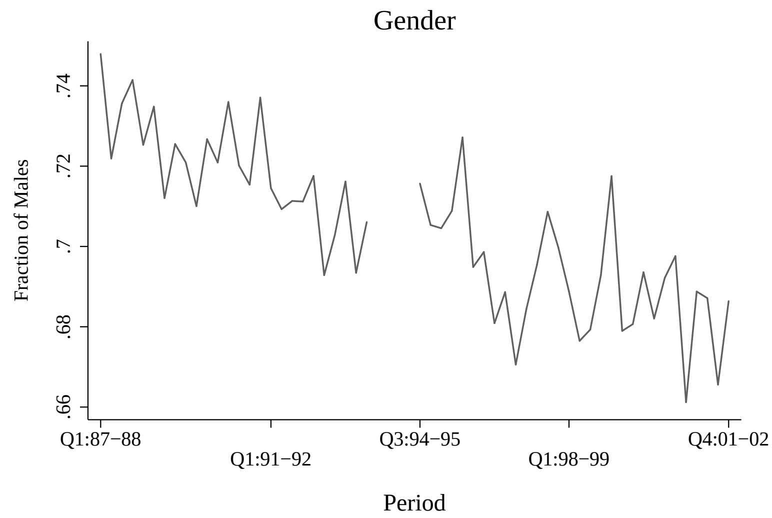


Figure 2.4: Fraction of Males

The reader will notice that (with the exception of the GDP graph which comes from National Accounts statistics), all the other graphs have a break in the middle, where no data is reported. This will happen with all the figures generated with the ENEU survey. The break comes from omitting those panels where the questionnaire changed, and for which no analysis was performed.

Figure 2.4 plots the fraction of males over time in the sample of study. This fraction is decreasing over time, due to the steady rise in the labor force participation of women.

Finally, the sectoral and regional composition of the sample are presented in Figure 2.5. Before analyzing this figure, it is important to clarify how the sector variables are constructed. In order to classify an individual as being in the formal or the informal sector, a question in the survey that asks whether the firm at which the individual worked last week had a name, or whether it was registered with the authorities, was used. An individual is considered to work in the informal sector if he reports that such firm did not have a name, nor it was registered with the authorities. To cross-check this statement, the individual must have not received any fringe benefit from this job, like health coverage, housing, social security, etc. In addition to that, if an individual was a self-declared informal street vendor, or worked at a firm with less than 6 employees that provided no fringe benefits at all, then he was also considered to be in the informal sector. The reason for this last classification choice is that some people might work at an informal firm that has a name, but no official registration. This is likely to be the case of individuals working at micro scale firms that provide no coverage to their workers. The classification into formal and informal sectors is further broken down by whether the individual is self-employed or a wage worker. There is evidence that wage workers and self-employed can differ dramatically in their characteristics (see for instance Maloney, 1999).

Figure 2.5, shows that approximately 70% of the individuals in the sample are formal wage workers, 20% are informal self-employed, around 8% are informal wage workers, and a very small remaining fraction are unemployed and formal self-employed. The low fraction of unemployed individuals in the sample is a known feature of the Mexican labor markets. In general, unemployment is low in Mexico, because there are no institutions that can cover a long search period for an individual who suddenly loses a job. Furthermore, the unemployment rate in this sub-sample of middle-aged workers seems to be lower than the unemployment rate of the overall urban population which is around 3.5%. Notice that all these proportions are fairly constant over time.

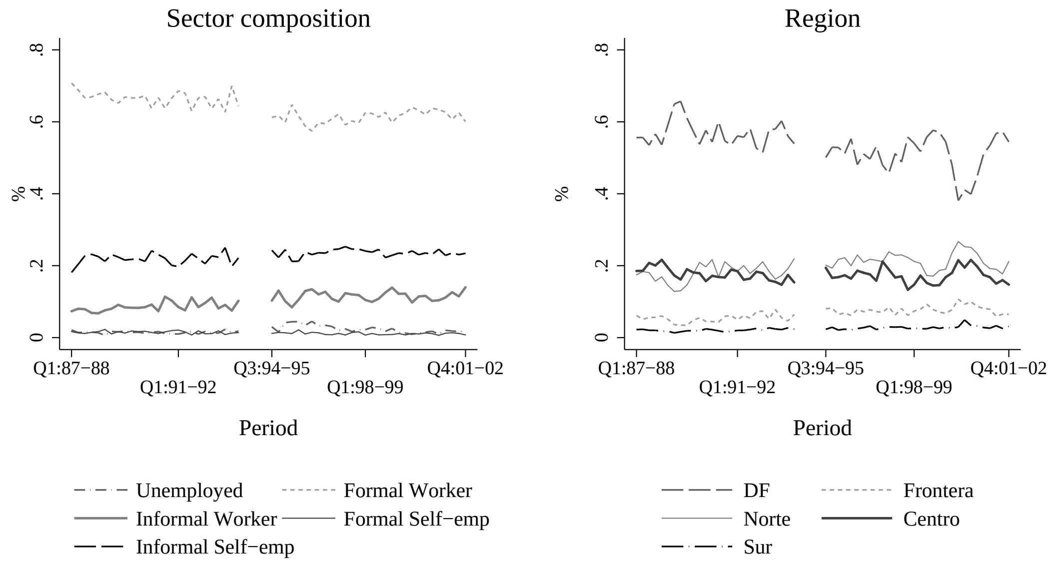


Figure 2.5: Sectoral and Regional Composition

Regarding the regional composition of the sample, most of the individuals included come from Mexico City (less than 60%), while the North and the Center regions in the country represent approximately 20% each. Finally, the US Border cities and the South region represent a small fraction of the sample. As previously mentioned, the ENEU expanded its geographic coverage over time, but this study focuses only on the present in the 1987 survey, in order to avoid confounding effects from expanding such coverage. The list of cities included under each region appears on the appendix at the end of the chapter, together with the number of observations contained in each panel.

2.4 Attrition and Non-response in the Survey

This section provides a descriptive look at the problem of missing individuals and non-reporting in the panel. Figure 2.6 displays the number of missing individuals in the panel after one year, as a fraction of the potential population of interest. It is evident from the graph that around half of the potential population is missing from the panel after one year. This is a high number and raises serious concerns about the representativeness of any mobility study performed with the ENEU. The bottom two graphs plot the reasons why individuals are not included in the sample. The list of possible reasons include attrition (i.e., disappearing from the sample in further re-interviews), mismatch according to variables of gender, age and education, missing earnings information, missing dwellings information,[7] missing sector information, and outliers in the earnings variables.[8]

From these graphs it is clear that the main reason for missing individuals from the sample is attrition. The fact that the ENEU tracks dwellings and not households explains part of this high attrition. It is important to note that the peak of attrition found around 1988-89 is of a different nature than the attrition present at other years. In these years it was not that an exceptionally high number of individuals were leaving their households, but rather that entire households were not matchable over the panel years. Perhaps there was an undocumented change in the areas surveyed in the panel, or there could be mistakes in the coding of household identification numbers. In any case, this extra peak in attrition is less

[7]Relevant only after the third quarter of 1994 when the dwelling questionnaire was introduced.

[8]An individual was considered to have earnings beyond the normal if for a single month it reported having earned more than 30,000 US Dollars, and reports much smaller amounts in other interviews.

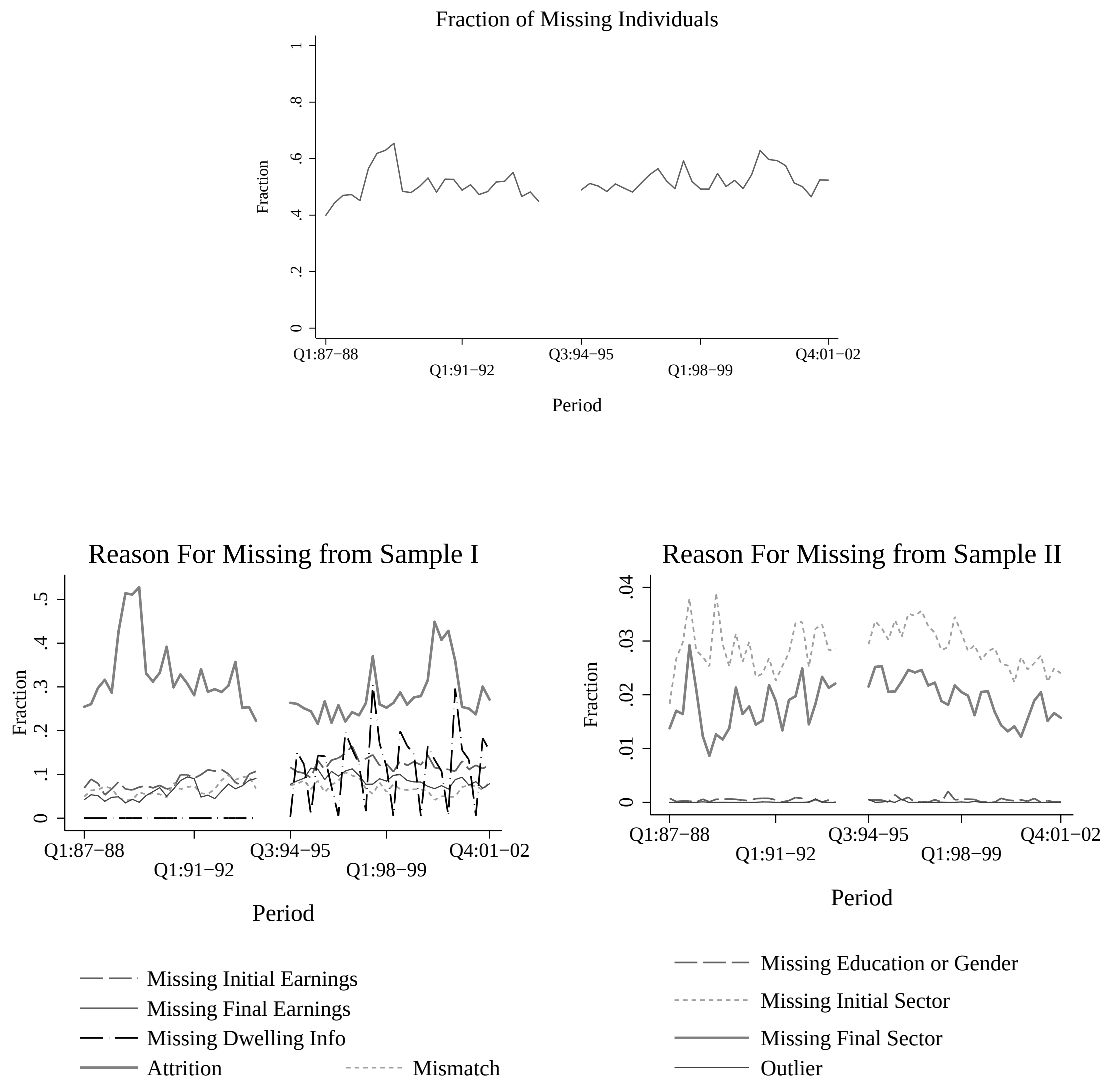
worrisome as it is unlikely to be driven by economic reasons, and it probably does not generate much bias in the estimates. Besides attrition, the other categories more relevant for the exclusion of individuals from the sample are missing earnings reports, missing dwelling characteristics, and mismatches.

The demographic characteristics for the missing individuals are presented in Figure 2.7.[9] This figure shows that there is a clear difference in the demographic characteristics of individuals who are missing because of attrition and mismatches, and the ones missing due to non-reporting of earnings. Overall, the latter are more educated, older, have a higher fraction of males, and have higher earnings.[10]

A comparison of Figure 2.7 with Figures 2.1-2.5 shows that the attritors do not differ much in their characteristics from the individuals included in the sample; however, non-reporters have substantial differences. This is a cause of concern, because if educated high-income individuals are not reporting their earnings when they experience positive mobility, then the mobility analysis in this dissertation could become biased.

[9] In the previous figure the categories were not exclusive, i.e., an individual could be mismatched and also not report earnings. For the present graph the categories *are* exclusive. This means that for an individual with multiple causes for being excluded from the sample, he will be first classified as an attritor (if he is one), and if he is not priority is then given to mismatch and finally to non reporting. Also notice that this figure only includes the main categories for missing.

[10] Most of the individuals who do not report earnings, do so at one point in time only. Hence, the earnings plotted for such individuals are either the earnings at the year before or at the year after, depending on when they refused to answer. Of course, there is a small fraction of individuals who do not report any of them.



Note: Categories are not mutally exclusive

Figure 2.6: Amount and Composition of Attrition

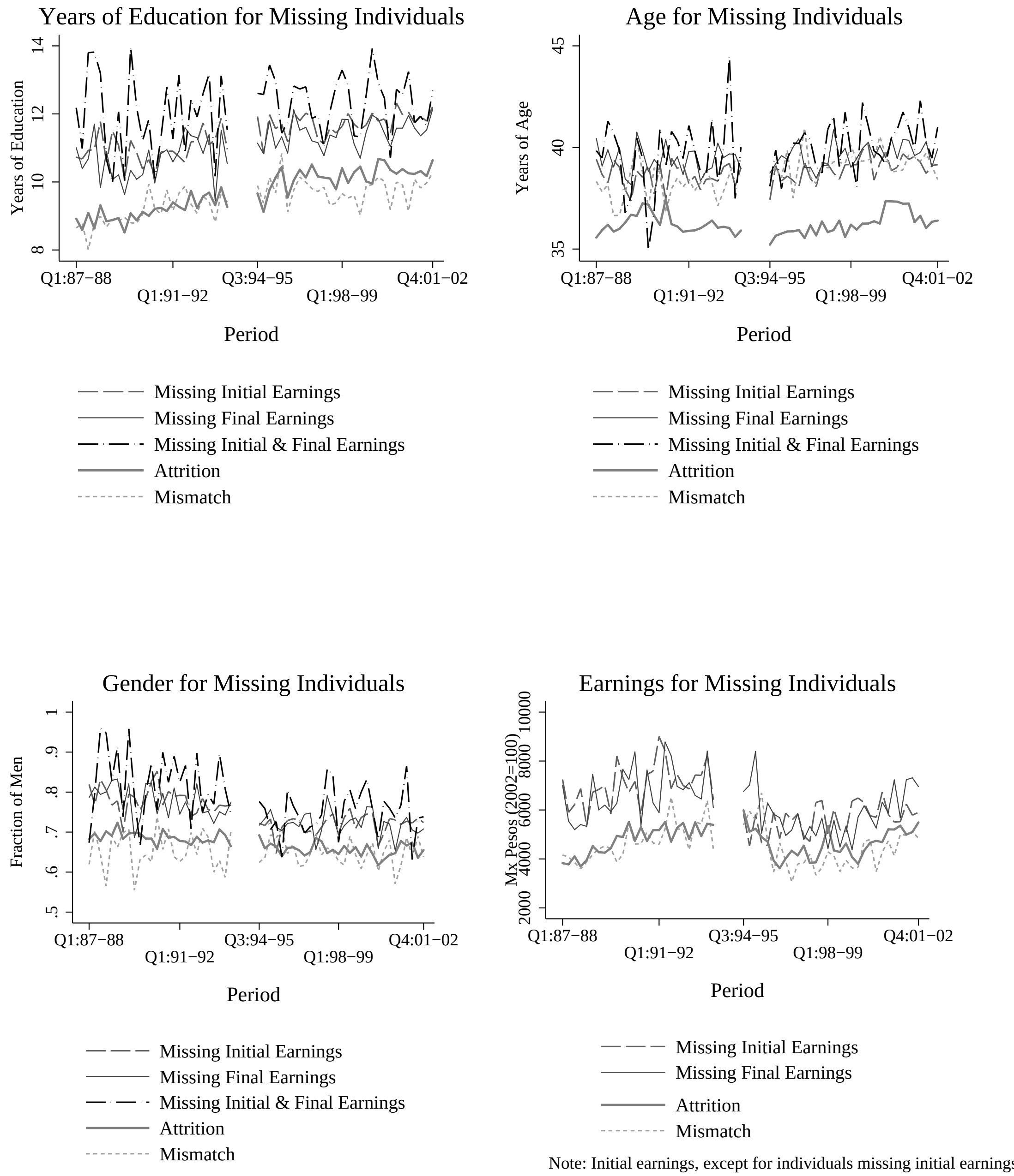


Figure 2.7: Demographic Characteristics of Missing Individuals

2.5 Final Remarks

This chapter introduced the survey used in this dissertation, as well some descriptive statistics for the sample that will be studied in the mobility analysis.

The sample under study is mostly composed of middle-aged males coming from Mexico city, working in the formal sector as wage workers with an average education at the high school level. However, it is also shown that some variables like education and gender composition present evolving trends: the education level of the sample increases, and there is also an increased participation of women in the labor force.

Regarding the macroeconomic evolution of the Mexican economy, it is shown that total output has risen, but the period under study contains a severe crisis that occurred in December 1994, and that created a sharp downturn in aggregate production.

The evolution of average earnings is growing in the period that goes from 1987 to 1994, but in the aftermath of the 1994 Peso crisis, earnings fell for the following 4 years, i.e., earnings continued falling even after total output had started growing back again. The most recent period from 1999 onwards witnessed an increase in average earnings, but this growth was not enough to match the pre-1994 levels. Finally, earnings inequality in the sample under study increased in the first half of the period (from 1987 to 1994) and remained steady in the following periods.

The amount of attrition and non-reporting of the earnings variable is quite large. While attritors have similar characteristics to the individuals that remained in the panel, this is not the case for the individuals who fail to report their earnings at some point in time. The latter are usually older and more educated than the average individual in the sample. Also, there is a higher fraction of males among the “non-reporters” and, whenever these individuals reported earnings, these were higher than the average earnings of individuals in the sample. For all these reasons, the earnings mobility results presented in chapters 3 and 4 should not be generalized to the whole urban population without further research on the impact of attrition on the estimates.

The next chapter contains the study of aggregate mobility trends for several groups of the population.

2.6 Appendix

2.6.1 Regions and their Cities

City	Region
Mexico City	Mexico City
Guadalajara	Center
León	
Puebla	
Orizaba	
Veracruz	
Chihuahua	North
Monterrey	
Tampico	
Torreón	
San Luis Potosí	
Mérida	Sur
Ciudad Juárez	US Border
Tijuana	
Matamoros	
Nuevo Laredo	

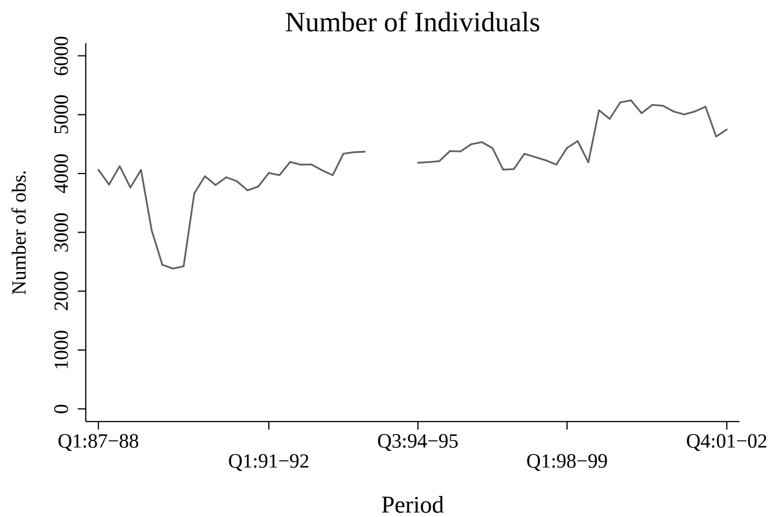


Figure 2.8: Number of Observations per Panel

Chapter 3

Aggregate Earnings Mobility

3.1 Introduction

This chapter presents the analysis of indices of aggregate earnings mobility. In particular, it focuses in studying two mobility concepts: *Directional mobility* and *Mobility as an equalizer of longer-term incomes*. As it will be argued below, these two concepts are particularly interesting from a normative point of view. Their analysis is also a good starting point in grasping the evolution of aggregate trends in earnings mobility for the Mexican economy.

This chapter is structured as follows. Section 3.2 motivates the chapter, while section 3.3 reviews the literature. The methodology utilized in this paper is explained in section 3.4, and results follow in sections 3.5-3.5.3. Section 3.6 concludes.

3.2 Motivation

The study of aggregate earnings mobility, also known as "macromobility", has a long tradition in Economics. Actually, it is with this approach that the mobility literature started in both Economics and Sociology. The reason for this lies in the fact that it was easier to calculate these indices when there were limited data and the computational resources to process these data were poor.

Aggregate mobility indices provide answers to the generic question "How much mobility of a certain type takes place in the economy?". This question is usually analyzed for different mobility concepts, and for different groups of the population.

Broadly speaking, the different aggregate mobility indices can be considered as quantifying 6 different mobility concepts. These are: Time dependence, Positional mobility, Share mobility, Symmetric mobility, Directional mobility and Mobility as an equalizer of longer-term incomes (for a detailed discussion of these concepts see Fields, 2004a).

Time dependence studies the mobility of how "one's current economic position is determined by one's position in the past" (Fields, 2001, p.107). Actually, Time dependence is concerned not only with the intertemporal relationship between economic *positions*, but also with the relationship between *levels* of income (or earnings).[1] Since the relationship between changes in earnings and initial earnings *levels* is studied in the following chapter at the individual level, the study of this concept is not pursued here.

The study of Positional mobility is concerned, as its name indicates, with the study of the changes in positions in the income distribution experienced by a group of economic agents. The unappealing side of this concept is that many interesting mobility processes can occur without properly implying any positional mobility. Consider for instance the hypothetical mobility process referred to in footnote 1, or another one in which everybody's income is increased by x pesos. In both cases there is no positional mobility, even when the agents in these economies experienced substantial changes in their incomes.

The concept of Share mobility is concerned with the study of changes in the shares of individual incomes as a proportion of the overall income level of the population. The problem with this approach is that a proportionate income change

[1] To appreciate the difference between these two cases consider the hypothetical situation in which \$1 was given to the poorest individual, \$2 to second poorest, and so on. Then, in this economy there would be no mobility in terms of *positions*, but the time dependence of levels of income would change, because people with different initial levels experienced different mobility patterns.

(i.e., where everybody's income is multiplied by a positive factor κ) would not be considered as mobility, even when the income levels might have changed significantly.

The first two mobility concepts (together with positional Time dependence) capture closely the idea of relative movements of incomes. Although, it is important to acknowledge that from a welfare perspective, individuals care not only about the changes in their absolute income levels, but also about the changes in their relative positions in society, it is not clear to the author that any of the indices used in the literature measure the relative mobility that the individual might actually care about.[2]

The only mobility concept of relative content that will be analyzed in this dissertation is Mobility as an equalizer of longer-term incomes. This concept answers the question "Does the mobility experienced in the economy help in equalizing the incomes of the individuals over time?". In particular, it captures one of the earliest concerns of the economists interested in mobility: whether the mobility process that takes place over time reduces the initial inequality that was observed in the economy (see Fields, 2004a, for a discussion of this concept and a new method to appraise it). If lifetime incomes are less unequally distributed than the initial ones, then it will be said that mobility equalized longer-term incomes. For the case of the present application, the data under study does not contain a measure of lifetime earnings (since the panel data used only tracks individuals for one year). However, this concept can still be applied to the Mexican case in order to assess to what extent the accumulated mobility over a year equalized the earnings of the

[2] In order to do that, it would be necessary to know to whom the individual is comparing herself, information which clearly is not collected in most of the surveys. For an interesting discussion of these issues see Frank (1992).

individuals initially observed.

Turning now to concepts of absolute mobility, the concept of Symmetric mobility measures how much incomes change in an economy without regard to whether these changes are gains or losses. Because of this insensitivity to the direction of the mobility experienced, this concept is not studied here. Instead, Directional mobility is studied. The concept of Directional mobility answers the question "What are the earnings gains and losses in the economy?". As its name indicates, this concept has the normative appealing feature that is sensitive to both the magnitude *and* the direction (gains or losses) of changes in income. If an individual gains an extra amount of income she will be considered to be better-off than before, independently of what happened to the incomes of the rest of the individuals in the economy. The opposite will be said in the case of an income loss.

3.3 Previous Research and Contribution

An overview of the literature on aggregate earnings mobility lies outside the scope of this section. A good introduction to the numerous indices developed in this literature and the mobility concepts they try to capture can be found in Fields and Ok (1999a) and Fields (2001). Instead, this section will discuss some of the references for the methodologies used in this chapter, as well as briefly overview some of the previous studies relevant to Mexico and other developing economies.

As previously mentioned, the questions of aggregate mobility in general take the form of "how much mobility is there," for a given mobility concept. The answers to such questions are usually provided by means of some index or other methods like transition matrices and stochastic dominance methods. Because calculating measures of aggregate mobility required less data than other mobility studies (like the ones dealt with in the following chapter), this area was tackled empirically early on. However, most of the research focused on measuring concepts of relative mobility; see Fields and Ok (1999a) for a survey.

The measurement of Directional mobility can be done in several ways. One simple way to approach it is by means of averages of earnings changes, both in levels and logs, for the the overall population and for several subgroups of it. The axiomatic properties of these indices are presented in Fields and Ok (1999b). Other possible methods include the comparison of the cumulative distribution functions (cdf's) of earnings changes in order to test for stochastic dominance (for an application to the mobility literature see Fields *et al.* (2002)), or the analysis of discrete transitions in-and-out of absolute income categories.

The first attempts trying to capture the concept of Mobility as an equalizer of longer-term incomes were made by Shorrocks (1978) and Chakravarty *et al.* (1985). The Shorrocks index has been criticized for having unappealing properties as to how it ranks mobility processes (see for instance Benabou and Ok, 2001; Fields, 2004a). An alternative measure of this mobility concept is the one offered by Fields (2004a). This measure has the properties that it is negative when mobility disequalizes incomes and positive when it equalizes them, relative to the initial distribution.[3]

Among recent empirical papers on mobility on developed countries presenting evidence of Directional mobility, one finds Buchinsky *et al.* (2003) and Fields (2004c) for France and the US. Among their conclusions are that Directional mobility presented a "saw-tooth pattern" in the US from the seventies to the early

[3]The Chakravarty-Dutta-Weymark (CDW) index is similar in nature to the Fields index if one replaces their 'hypothetical' income distribution path with the distribution at the initial period.

nineties, while in France it first rose considerably and then fell in the late seventies, stabilizing thereafter at a lower level.[4] When looking at mobility by population groups in France, the authors find that there are no significant differences by gender, and that individuals with higher education experienced higher positive mobility.

For developing countries the stories vary depending on the economic evolution of each country, and many times evidence on Directional mobility is presented only indirectly. This means that specific measures of this concept are not systematically calculated, but rather presented as a complement to other topics studied. An example are the numerous studies presenting evidence of the transitions in-and-out of (absolute) poverty. Clearly, such evidence provides results on Directional mobility. However, limiting the analysis to a discretization of the data (above and below the poverty line) leaves unaddressed many of the details that could be learned if the mobility across the whole distribution was studied.

Among the studies in Latin America that explicitly address Directional mobility one can find Herrera (1999) and Glewwe and Hall (1998) for Peru, Freije (2001) for Venezuela, and Sánchez-Puerta (2005) for Argentina. For Peru, there was positive mobility in the late eighties. Among the factors positively correlated with this upward mobility were the individual's education level and age. For Venezuela, a trend of impoverishment is found through the eighties and nineties. Households with higher human capital endowments (and hence usually with higher income too) tend to experience larger gains and losses than the rest of the population, but these changes are proportional to their initial income levels. Finally, for Argentina there

[4]Further analysis for the US presented in Fields (2001) by means of stochastic dominance analysis shows that in the 1980's more people lost more dollars than in the previous decades, but the winners, won more than before.

was a falling trend in earnings over the past decade, and age, gender and education do not seem to play a major role in explaining this mobility, except during the crisis of early 2000, when higher education was negatively associated with earnings changes and women fared slightly better than men. Of course, it is crucial to keep in mind that these results only describe the mobility experienced by subgroups of the population and do not constitute evidence of causal determination.[5]

The empirical evidence on Mobility as an equalizer of longer-term incomes is more conclusive in the sense that in general, both for developed and developing nations, mobility equalizes longer-term incomes. Either by means of explicit indices or by simple comparisons of inequality indices calculated over lifetime incomes versus their cross-sectional counterparts, this finding appears to be consistent for most of the studies.

Only a couple of papers have studied aggregate mobility issues for Mexico. The studies by Wodon (2001) and Yitzhaki and Wodon (2002) study *time-dependence* in the ranks of the individuals, the first comparing urban Mexico and Argentina, the second in rural Mexico. The first paper studies how one particular measure of time-dependence in ranks -the Gini Index of Mobility- evolves in both countries over different points of the business cycle. The study uses the same database as the present dissertation and covers from the late eighties to the mid-nineties. Among its main findings are that in Mexico there is less time-dependence during growth periods, while the opposite occurs for the case of Argentina. According to the author Mexican labor markets adjusted to negative macroeconomic shocks through price adjustments (i.e., wage cuts), while in Argentina they adjusted by

[5] Another grouping generally explored when analyzing Directional mobility is the quintile of the initial income distribution. Results pertaining to this analysis are commented in the next chapter.

means of changes in quantities (rise in unemployment). Since labor cuts are likely to lead to more re-ranking of individuals along the earnings distribution than wage cuts, Argentina experienced less time-dependence in ranks during the downturns of its economy. Another interesting finding is that young uneducated workers experience less time-dependence than the rest of the population.

The paper by Yitzhaki and Wodon (2002) uses a dataset collected by the World Bank and the Ministry of Agrarian Reform. This dataset was related to the rural subsidies program PROCAMPO. The study was conducted in rural areas in Mexico in 1994 and 1997. This mobility study analyzes time-dependence in ranks using again the Gini Index of Mobility for four welfare measures: Per-capita income, per-capita land owned, per-capita land cultivated and PROCAMPO transfers. Among the findings of these authors are that, in general, time-dependence in ranks is quite high in these rural samples, meaning that individuals preserve their ranks over time. They also find that the time-dependence is smaller with land measures as proxies of welfare than with per-capita income. Finally, they report that the PROCAMPO transfers had the effect of providing a limited re-ranking in the distribution.

As previously mentioned, both of these papers focus on time-dependence in ranks only, hence the evidence on aggregate Directional mobility and Mobility as an equalizer of longer-term earnings presented in this dissertation is new to the literature. Also, the decomposition of the measure of Mobility as an equalizer of longer-term incomes in between and within-group components is, as far as I know, new in the literature.

3.4 Methodology

The measurement of Directional mobility can be done in several ways. Two popular methods used in the literature are the estimation of the average gains (and losses) experienced by a group of individuals, and the use of stochastic dominance analysis over the distributions of income changes. Since in this dissertation Directional mobility is quantified for several subpopulation groups and for 56 overlapping panels, the first method is preferred for its compactness. One disadvantage of taking this route is that, by taking averages, the amount of mobility could be understated, because positive and negative changes would cancel each other out.[6] In principle, performing this analysis for different subgroups of the population should reduce this problem, but does not completely eliminate it. It is important to keep this caveat in mind for the rest of this chapter.

Denoting by y_{it} the monthly earnings of individual i at time t, the two indices of Directional mobility estimated are the average earnings changes, i.e.,

$$\overline{\Delta y_s} = \frac{1}{N_s}\sum_{i=1}^{N_s}(y_{it} - y_{it-1}) \tag{3.1}$$

and the average log-earnings changes, i.e.,

$$\bar{m}_s = \frac{1}{N_s}\sum_{i=1}^{N_s}(\ln y_{it} - \ln y_{it-1}) \tag{3.2}$$

where $i \in s$, and s is an arbitrary subgroup of the population of size N_s (it might be the overall population as well).[7]

[6] The comparison of distributions of earnings changes does not suffer from this problem.

[7] For the case of (3.2) the calculations are performed only for individuals working and reporting positive earnings.

The reason for analyzing average log-earnings changes, in addition to changes in levels, is that the former measure gives higher weight to the earnings mobility of individuals at the bottom of the earnings distribution. In general, the initially poor will experience less downward mobility simply because they had a smaller initial income. Taking logarithms and their change is a partial solution to this problem, because it gives a higher weight to the earnings changes of the poor, and approximates proportional changes in earnings. The index $\bar{m}_s$ has been axiomatized by Fields and Ok (1999b).

Regarding the concept of Mobility as an equalizer of longer-term incomes the index used is the one proposed by Fields (2004a). This index has the properties of being: (i) decreasing in the inequality of long-term incomes, (ii) increasing in the inequality of initial incomes, and (iii) equal to zero if these two inequality measures are equal, i.e., if mobility did not bring a change in the inequality in the economy over the long-run.

The particular functional form of this index is

$$\bar{P} = 1 - \frac{I(\bar{y})}{I(y_{\cdot t-1})} \tag{3.3}$$

where $I(\cdot)$ is a measure of inequality, $\bar{y}$ is a vector of individual average earnings (the average being taken over time for each individual), and $y_{\cdot t-1}$ is the vector of initial earnings. If mobility (dis)equalizes longer-term earnings, the index will be (negative)positive, otherwise it will be zero. Other indices have been proposed in the literature to measure this mobility concept. In particular the index proposed by Shorrocks (1978) has been commonly applied however, as was previously mentioned recent research has found unappealing properties in this index and hence it won't be used here.

The concept of Mobility as an equalizer of longer-term incomes is often considered as being equivalent to a reduction in inequality in the economy (as measured by the changes in an inequality index). A simple example borrowed from Fields (2004a) will demonstrate that this is not the case. Consider a 2-person economy in which the incomes of the individuals conforming it are represented by a vector (y_1, y_2). The first entry in the vector always corresponds to the income of person 1, and the second entry to the one of person 2. Suppose furthermore that the specific initial incomes are given by (1,5). Now consider two scenarios,

	Period 1	Period 2
Scenario I:	(1,5)	(1,6)
Scenario II:	(1,5)	(6,1)

Under both scenarios inequality grew (the Gini index in Period 1 is 0.33, while in Period 2 equals 0.35), however Scenario I disequalized longer-term incomes (the "longer-term" incomes being just the averages over time for each individual). Indeed, the Gini index for the vector of longer-term incomes (1,5.5), is 0.34. On the other hand, Scenario II equalized longer-term incomes, since they are (3.5,3), with an associated Gini index of 0.03. This example demonstrates that inequality, when measured cross-sectionally, can grow and still the mobility process can equalize longer-term incomes.

The inequality measure used in this dissertation for the P-index will be the Generalized Entropy index with parameter $\alpha = 2$, or GE(2) (this equals half the squared coefficient of variation). The reason for selecting this index (instead of the more popular Gini index) is that it is Lorenz-consistent, it allows for earnings to take values equal to zero, and unlike the Gini, it is decomposable in inequality between groups and inequality within groups for some groups of the population.[8]

An inequality index that is decomposable into between and within-group inequality components can be written as

$$I(y) = \omega_w I_w(y) + \omega_b I_b(y)$$

where y is the vector of incomes, I_w and I_b are the within and between-group inequality indices, and ω_w and ω_b are their respective weights (which can equal 1, as in the case of the GE(2)).

In this dissertation, a similar decomposition is proposed for the P-index. In particular, a simple algebraic manipulation shows that the P-index can be written as

$$P = \kappa_w P_w + \kappa_b P_b$$

where $P_w = 1 - (I_w(\bar{y})/I_w(y_{\cdot t-1}))$ and $P_b = 1 - (I_b(\bar{y})/I_b(y_{\cdot t-1}))$ are the indices capturing whether mobility helped in equalizing longer-term incomes within and between groups of the population respectively. The κ-weights equal $\kappa_b = \omega_b I_b(y_{\cdot t-1})/I(y_{\cdot t-1})$ and $\kappa_w = \omega_w I_w(y_{\cdot t-1})/I(y_{\cdot t-1})$. In the case of an index having weights $\omega_w = \omega_b = 1$ (such as the GE(2)) the κ-weights are just the between and within-group base period inequalities, as fractions of the total initial inequality. This decomposition will be useful in answering the questions "Does mobility equalize earnings within groups over time?" and "Does mobility equalize earnings between groups over time?" when the analysis by groups of the population is performed.

Having presented the main indices used in this chapter to capture aggregate earnings mobility, the next section presents the results obtained for urban Mexico.

[8]The decomposition of the Gini contains some residuals that are hard to interpret.

3.5 Results

3.5.1 Initial advantage by subpopulation group

Before presenting the results for the different mobility measures previously described, this section provides evidence on which subgroups of the population experience higher initial earnings, something considered to be a measure of initial advantage.

Figure 3.1 contains the initial average earnings profiles for individuals grouped by quintiles of the initial earnings distribution. Although in this case the ordering of who earns more is trivially determined by construction, the figure illustrates how big are the differences in earnings across different points in the earnings distribution. In particular, this graph shows the large differential in earnings between the individuals in the top quintile and the rest of the population. While the earnings of the first 4 quintiles are between 1000 and 5000 Mx pesos a month, the earnings of the individuals in the 5th quintile is usually above the 10000 Mx pesos. The picture on the right shows the F-statistic for the hypothesis test that earnings are equal between groups, together with the critical value of the F distribution at the 95-percentile. It is clear from the graph that this hypothesis is rejected.

The second group analyzed is age groups. For this purpose the population is divided into three age groups. The first group contains individuals between 25 and 36 years of age, the second, individuals between 37 and 48 years, and the last one contains the individuals with age between 49 and 60 years. Figure 3.2 shows there is no clear advantage in terms of initial earnings for any of these age groups. If anything, the middle-aged group (going from 37 to 48 years) seems to have slightly higher earnings than the youngest one. The F-test statistic confirms the intuition

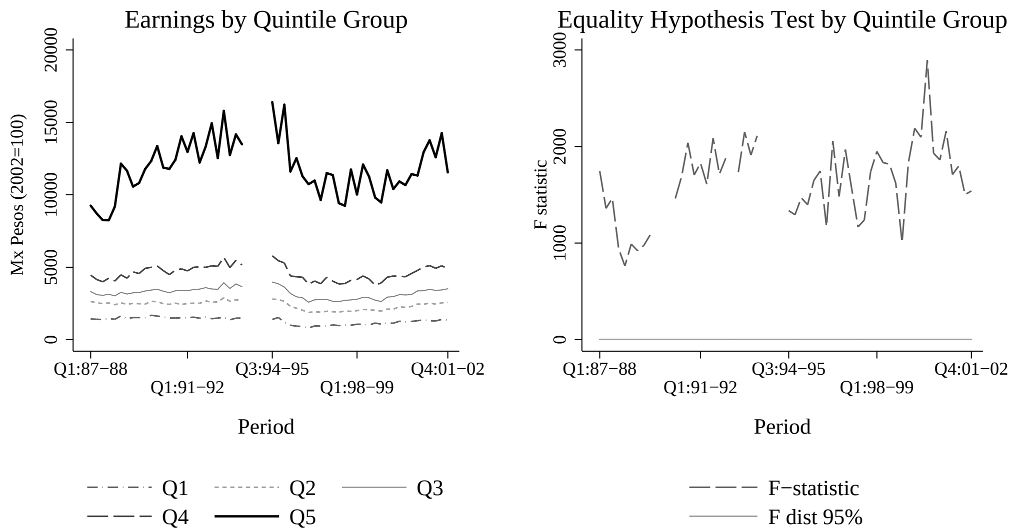


Figure 3.1: Initial Earnings by Quintile Group

that earnings are most of the times roughly equal among age groups.

For the analysis of education groups, the population was divided into individuals with Elementary (0 to 6 years of schooling), Secondary (7 to 12 yrs.) and Higher education (12 or more yrs. of education). The results in Figure 3.3 present a clear ranking of individuals, where individuals with more education have higher initial earnings. The individuals with higher education earn at least twice as much as the ones with secondary education, and these in turn earn about 1.3 times what the individuals with elementary education do. All these differences are statistically significant, as the second graph shows.

Figure 3.4 presents the results for gender categories. The initial advantage in terms of earnings favors male individuals which earn about 1500 Mx Pesos more than their female counterparts. The hypothesis test for equality across groups

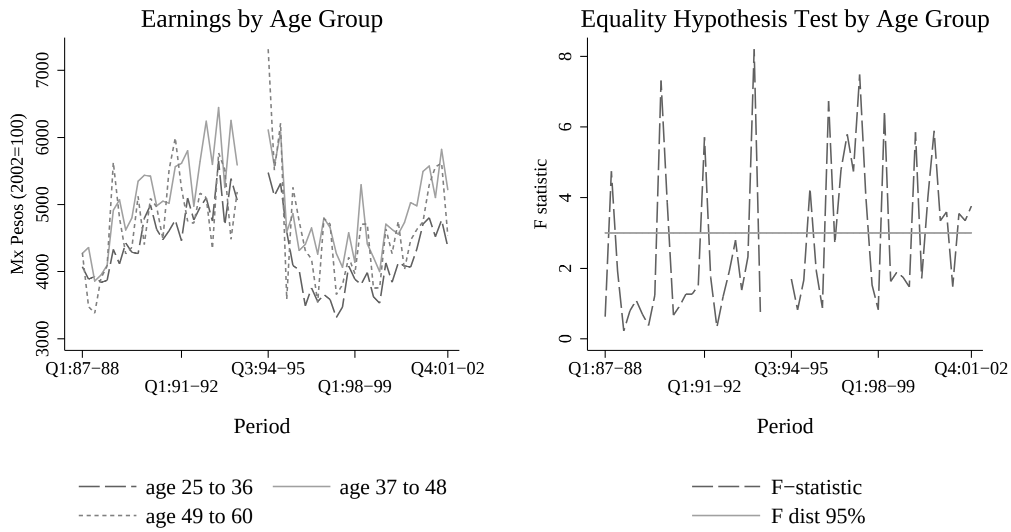


Figure 3.2: Initial Earnings by Age Group

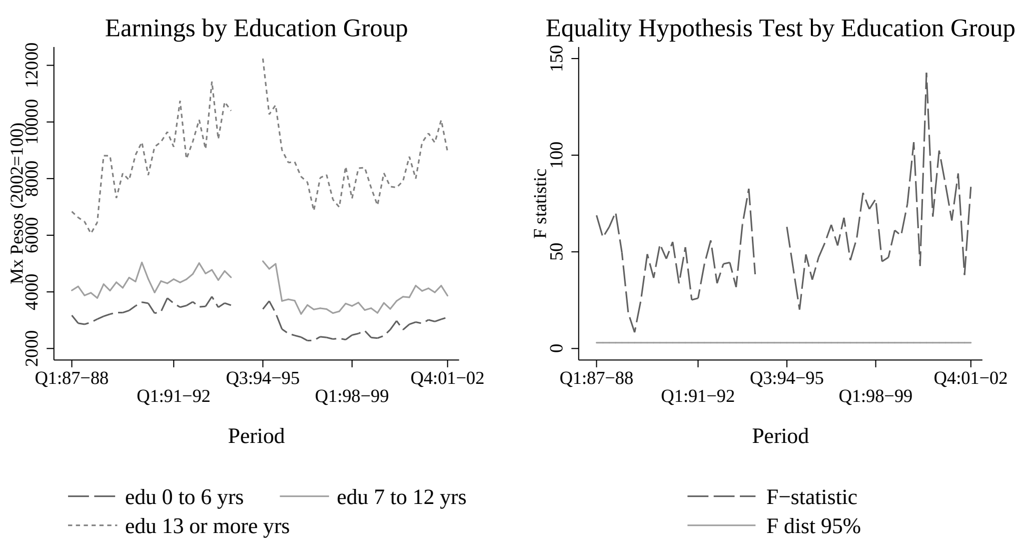


Figure 3.3: Initial Earnings by Education Group

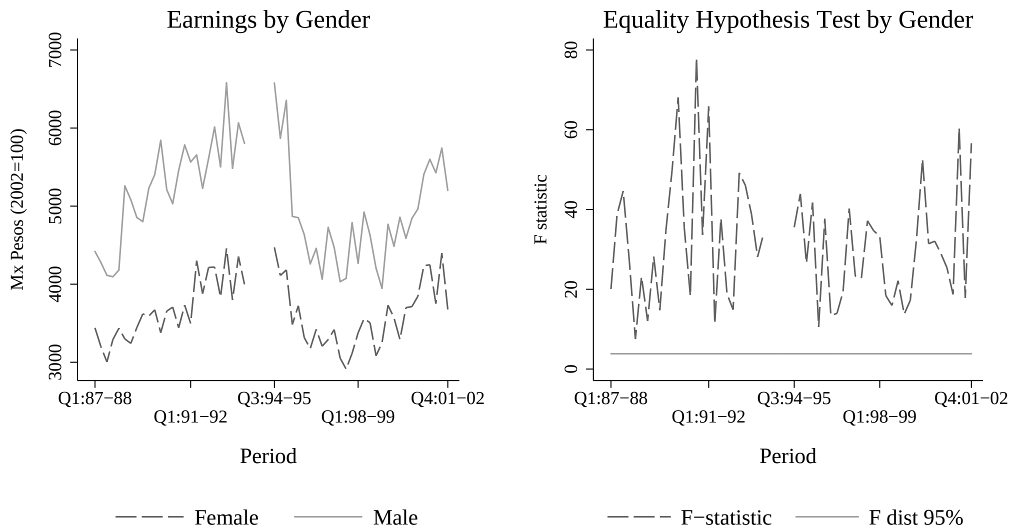


Figure 3.4: Initial Earnings by Gender

confirms that this difference is statistically significant for all the periods in the sample.

Regarding the initial earnings of individuals in different sectors, Figure 3.5 shows that the earnings of formal self-employed are substantially higher than those of anyone else. The graph also shows that the earnings of formal wage workers and informal self-employed are roughly the same. At the bottom of the earnings distribution stand the informal wage workers which earn almost half of what their informal self-employed counterparts do, confirming the intuition that they are the worst-off individuals among the sector categories. Again the hypothesis test of earnings equality across groups confirms that these differences are statistically significant.

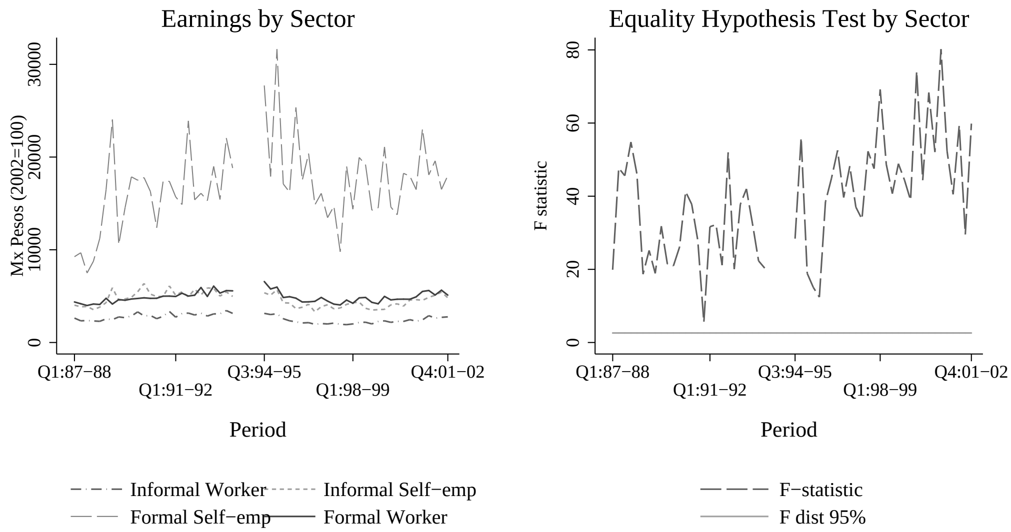


Figure 3.5: Initial Earnings by Sector

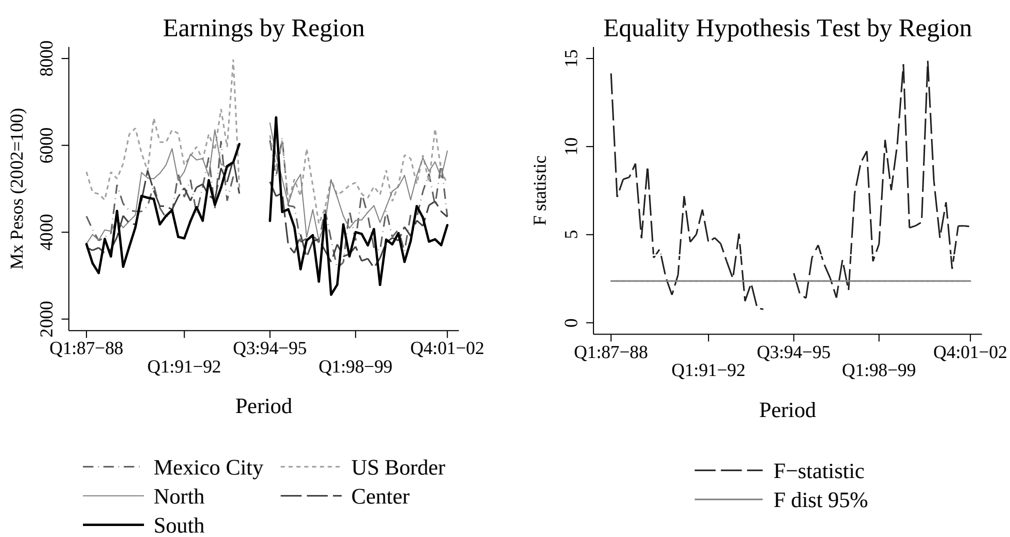


Figure 3.6: Initial Earnings by Region

Finally, the initial earnings by region plotted in Figure 3.6 show that individuals located in cities along the US Border have larger earnings on average, while individuals living in the South present usually the lowest earnings. The hypothesis of equality of earnings among regions is rejected most of the times.

3.5.2 Directional Mobility

Having presented which subgroups of the population have a higher initial advantage in terms of average earnings, this section proceeds to present the results on the indices of Directional mobility for the whole sample, and for the previously analyzed groups. Again, the measures analyzed are average earnings mobility in levels and logarithms.

The average mobility for all the individuals in the sample, both in levels and logarithms is plotted in Figure 3.7.[9] In this figure it is possible to see that the average mobility during the first half of the sample was most of the times zero, with the exception of moderate growth in average earnings and log-earnings during the late eighties. After the 1994 Peso crisis there was a strong downfall in earnings, and positive mobility occurred again only in the early 2000. Finally, the last years on the sample show small negative mobility in levels.

Concerning the earnings mobility for individuals at different quintiles of the initial earnings distribution Figure 3.8 shows that the higher the initial advantage, the higher is the loss suffered (or the smallest the gain). This conclusion holds independently of whether one looks at absolute earnings changes or proportional earnings changes (as approximated by the change in log-earnings). Individuals in

[9]This figure also includes 95% confidence intervals to indicate when these estimates are statistically different from zero.

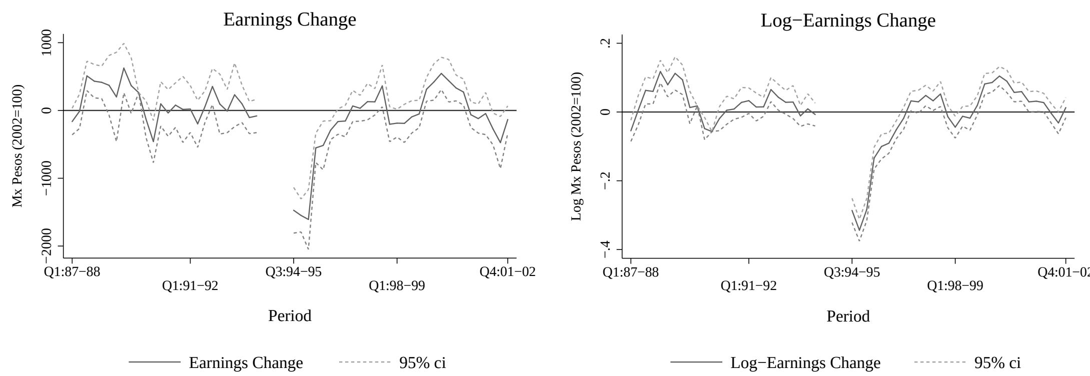


Figure 3.7: Average Earnings Mobility

the higher quintile experience considerable earnings losses over a year, even after accounting for the fact that they started at a higher position and hence had literally "more to lose". The result that individuals in the lowest initial quintile always experience the highest gains is not driven by the fact that this group contains unemployed individuals in the base period. This is because the unemployed are a small fraction of the sample.[10] All these findings provide support for the fact that there is some mean reversion in the earnings of the individuals in the short run, a topic that will be studied more carefully in the next chapter.

When looking at the earnings mobility by age and education groups in Figures 3.9 and 3.10, no clear pattern seems to emerge. In the case of age, the earnings mobility looks the same for the different subgroups, and the hypothesis tests of equality of mobility by groups fail to reject this hypothesis most of the times. The same occurs for the education groups considered. The earnings mobility in levels of the most educated individuals fluctuates more, but the graph of log-earnings changes confirms that these higher fluctuations are due only to the fact that this

[10] In fact dropping the unemployed from this analysis, leaves the conclusions unaltered.

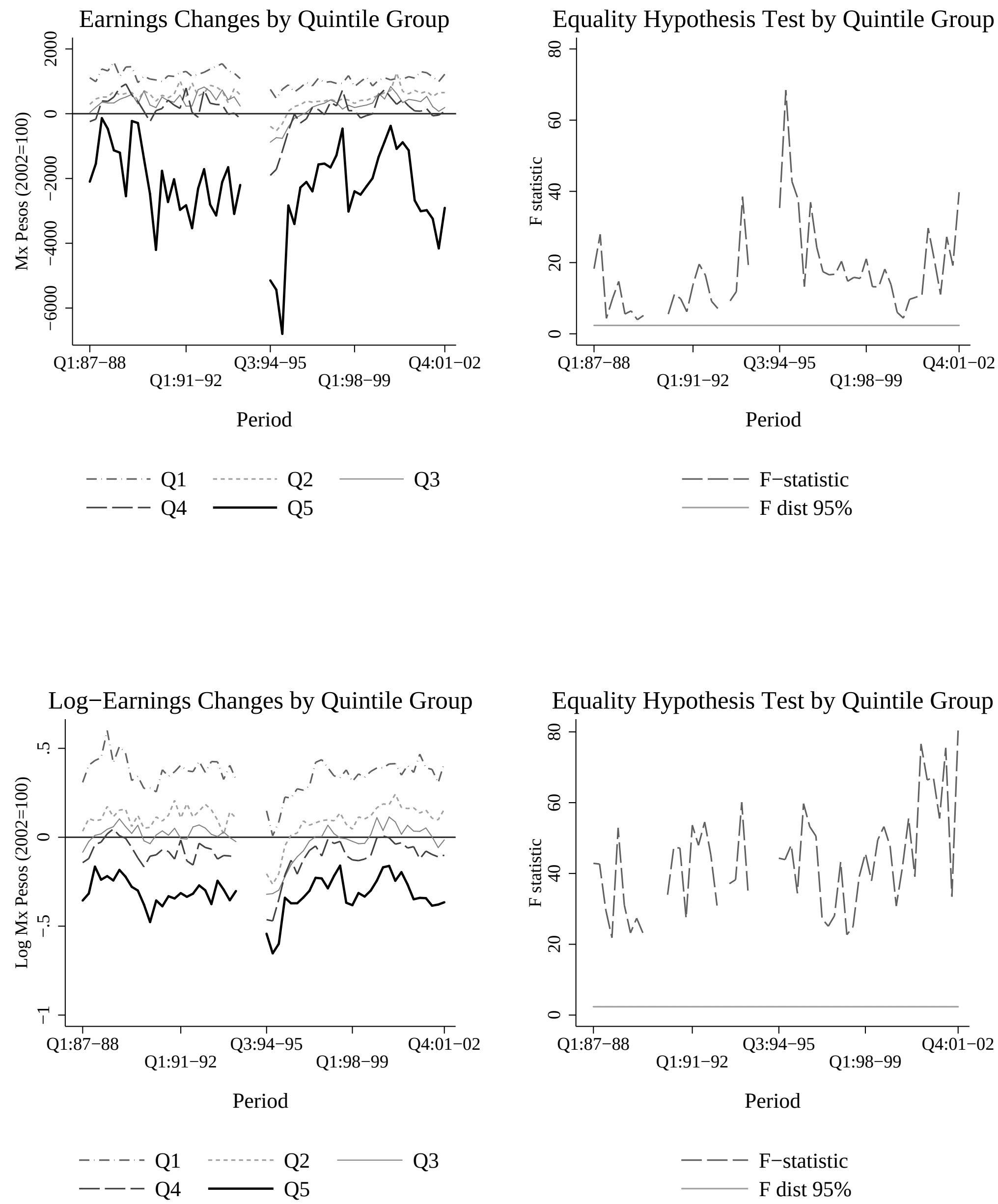


Figure 3.8: Average Mobility by Quintile Group

group has a higher level of earnings. In any case, the hypothesis tests fail to reject most of the times the equality of mobility for the different education groups. In other words higher levels of education are not associated with more positive mobility.

Regarding the earnings changes by gender one can observe that the earnings mobility experienced by men and women is pretty similar, with the exception of the aftermath of the 1994 Peso crisis when men experienced significantly higher losses in absolute levels. However, as the logarithms graph shows, these losses were proportional to the higher earnings men had before the crisis started.

Figure 3.12 shows, no clear mobility pattern emerges by region, neither in levels nor in logarithms. In both cases the hypothesis tests of equality of mobility patterns by region fail to reject such hypotheses.

The mobility analysis by sector shows that the unemployed individuals experienced the largest gains (since most of them found jobs), while the formal self-employed (who were the initially most advantaged in terms of earnings) experienced the largest losses. This indicates that part of the large initial advantage of the formal self-employed vanished after a year. The informal self-employed and the formal wage workers experienced on average the same mobility. The analysis of the log-mobility patterns confirms the results found in levels, and adds to the picture the fact that informal wage workers (who usually have the lowest earnings) are the ones who experience the largest proportional gains. The hypotheses tests performed reject the equality of mobility between sectors most of the times.[11]

The study of Directional mobility by sector can be refined by showing the

[11] It is important to remark that this hypothesis test for the mobility in levels excludes the unemployed. Including them would lead to over-rejection of the equality hypothesis, since this group never experiences negative mobility.

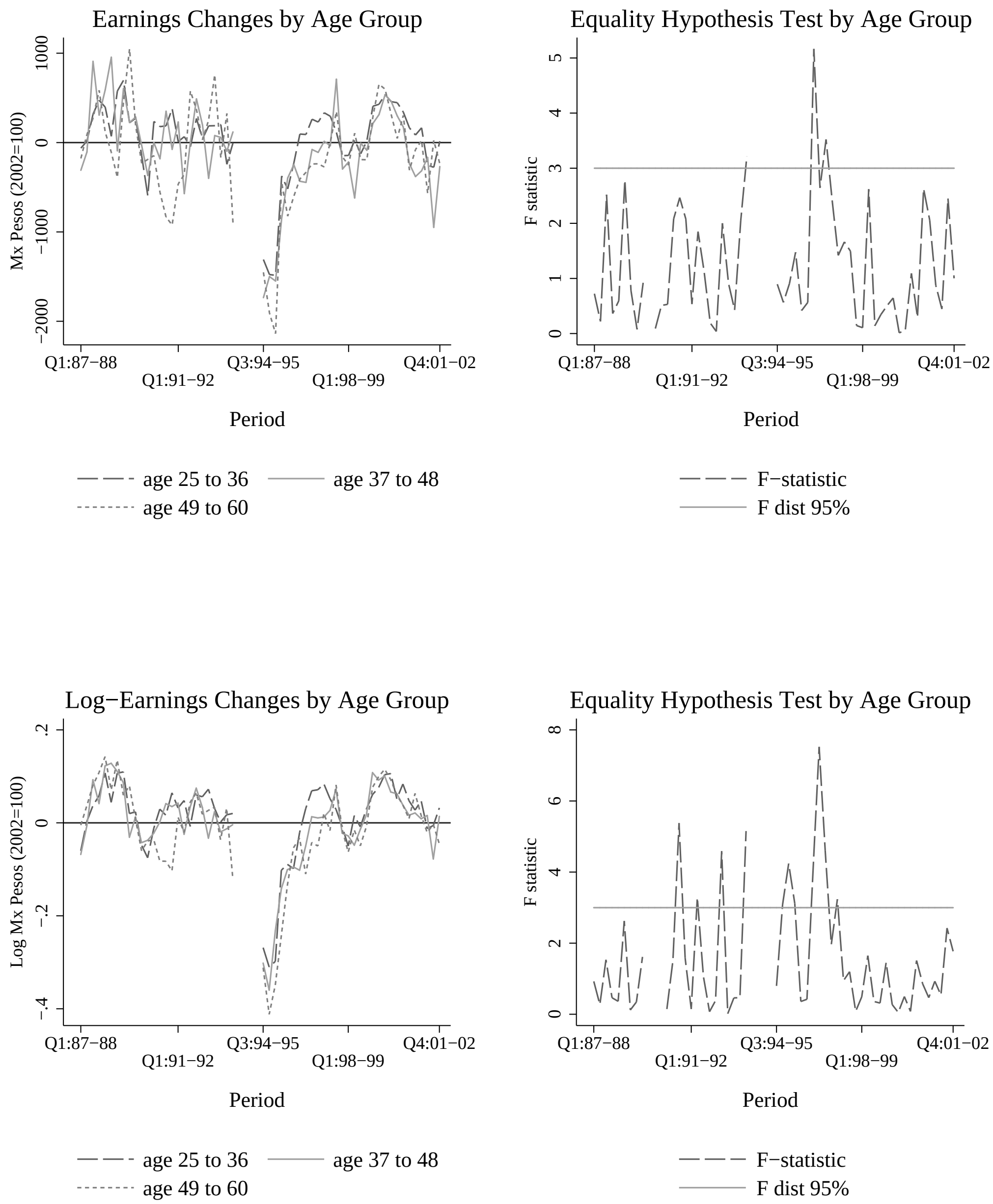


Figure 3.9: Average Mobility by Age Group

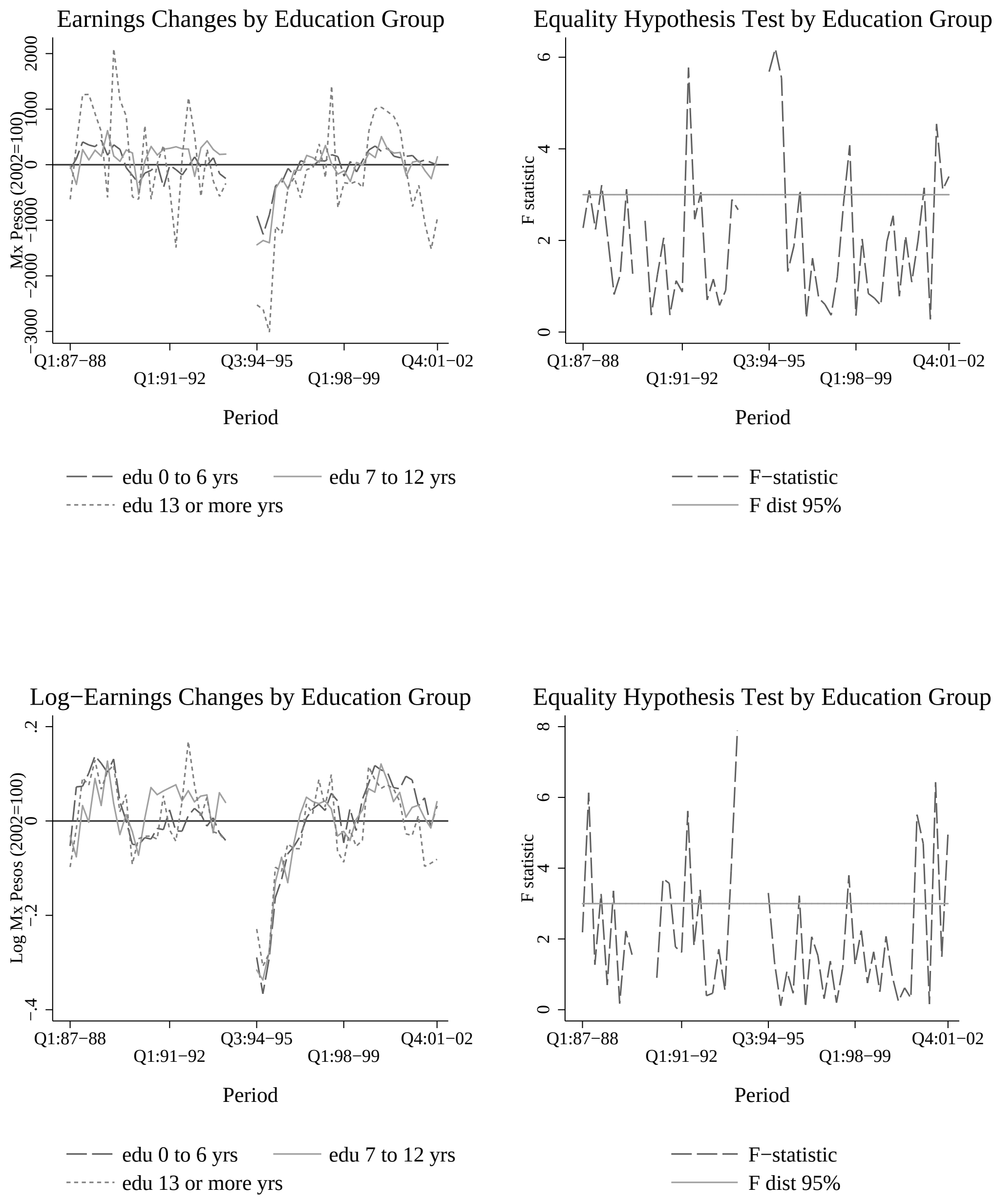


Figure 3.10: Average Mobility by Education Group

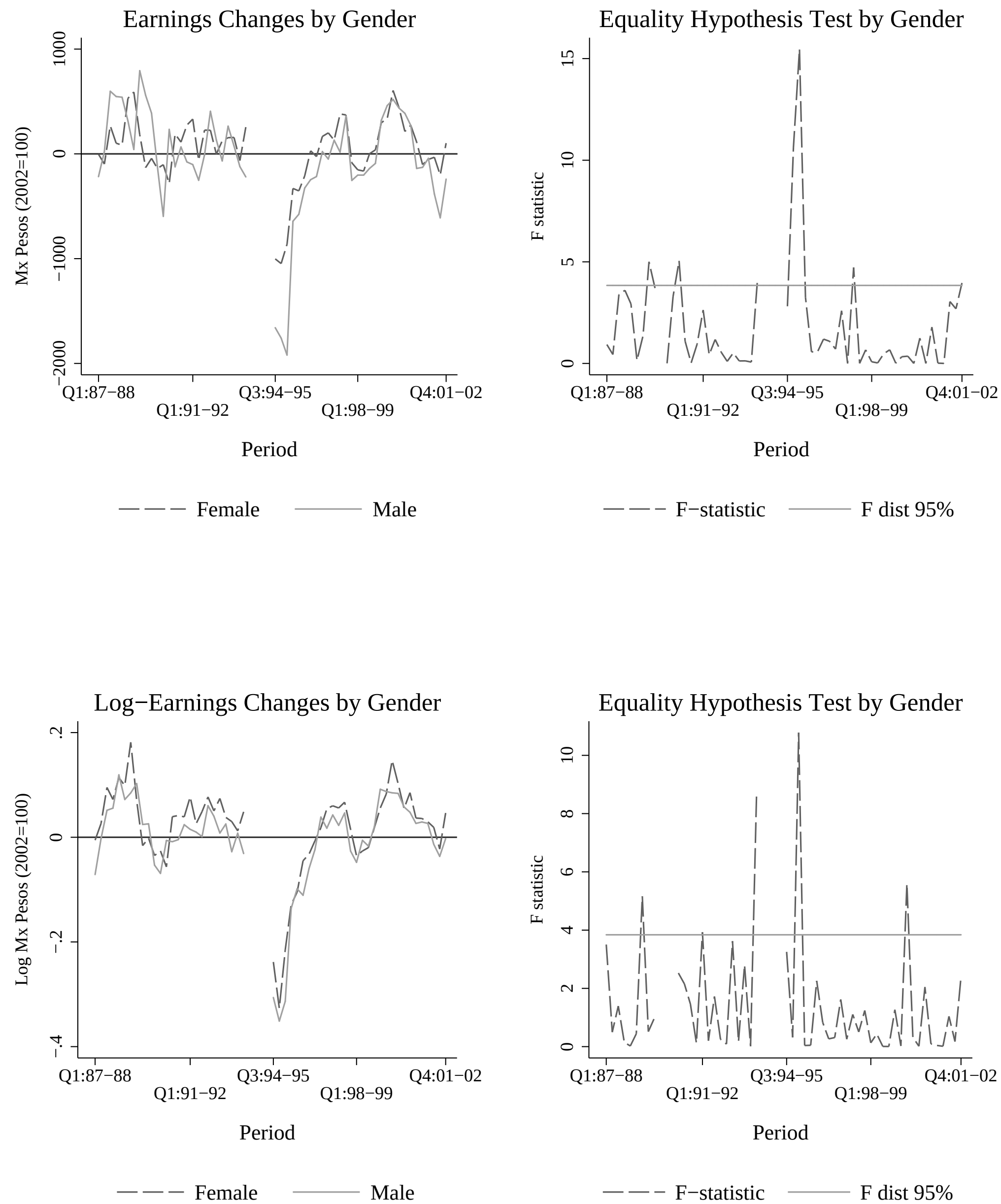


Figure 3.11: Average Mobility by Gender

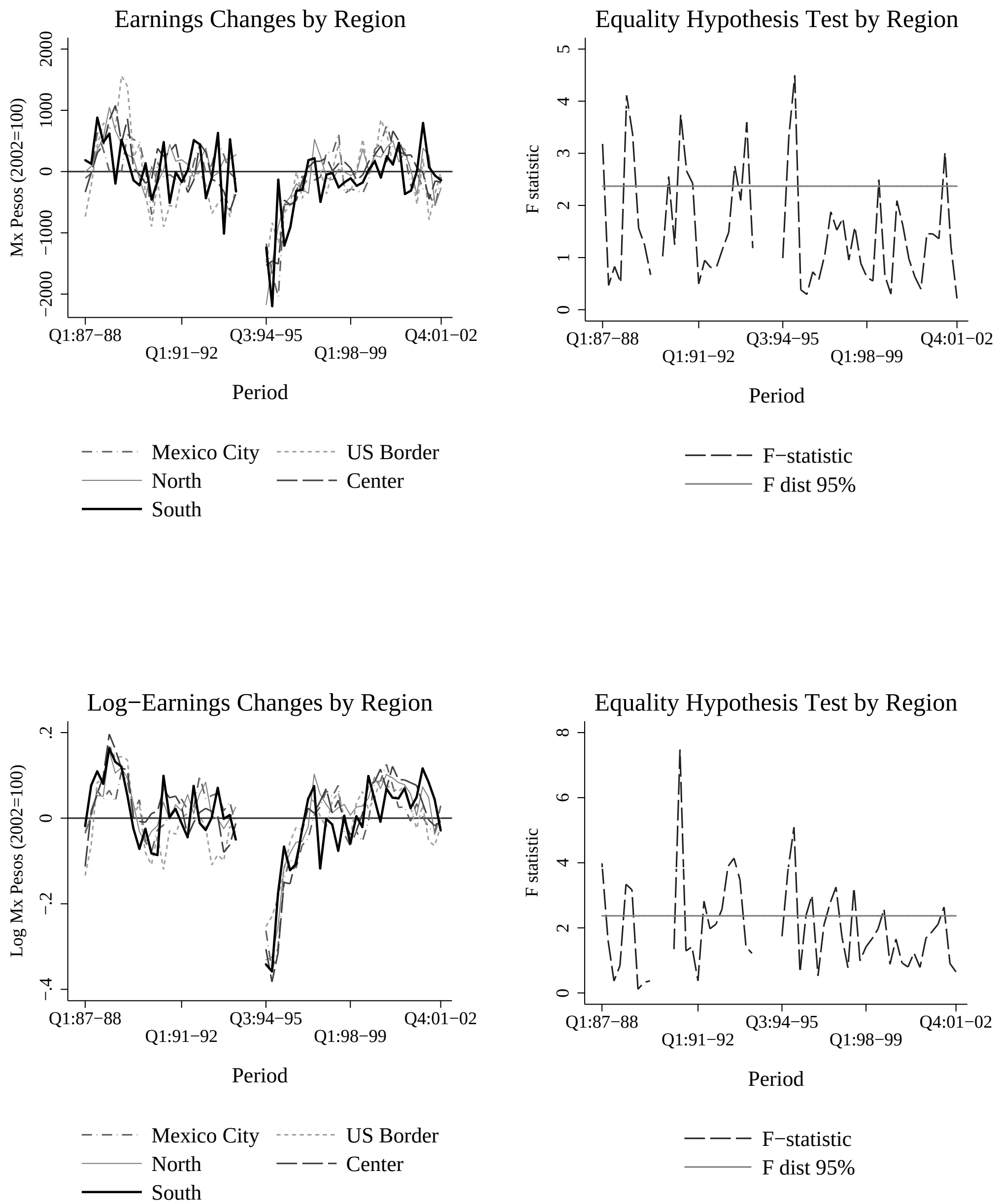


Figure 3.12: Average Mobility by Region

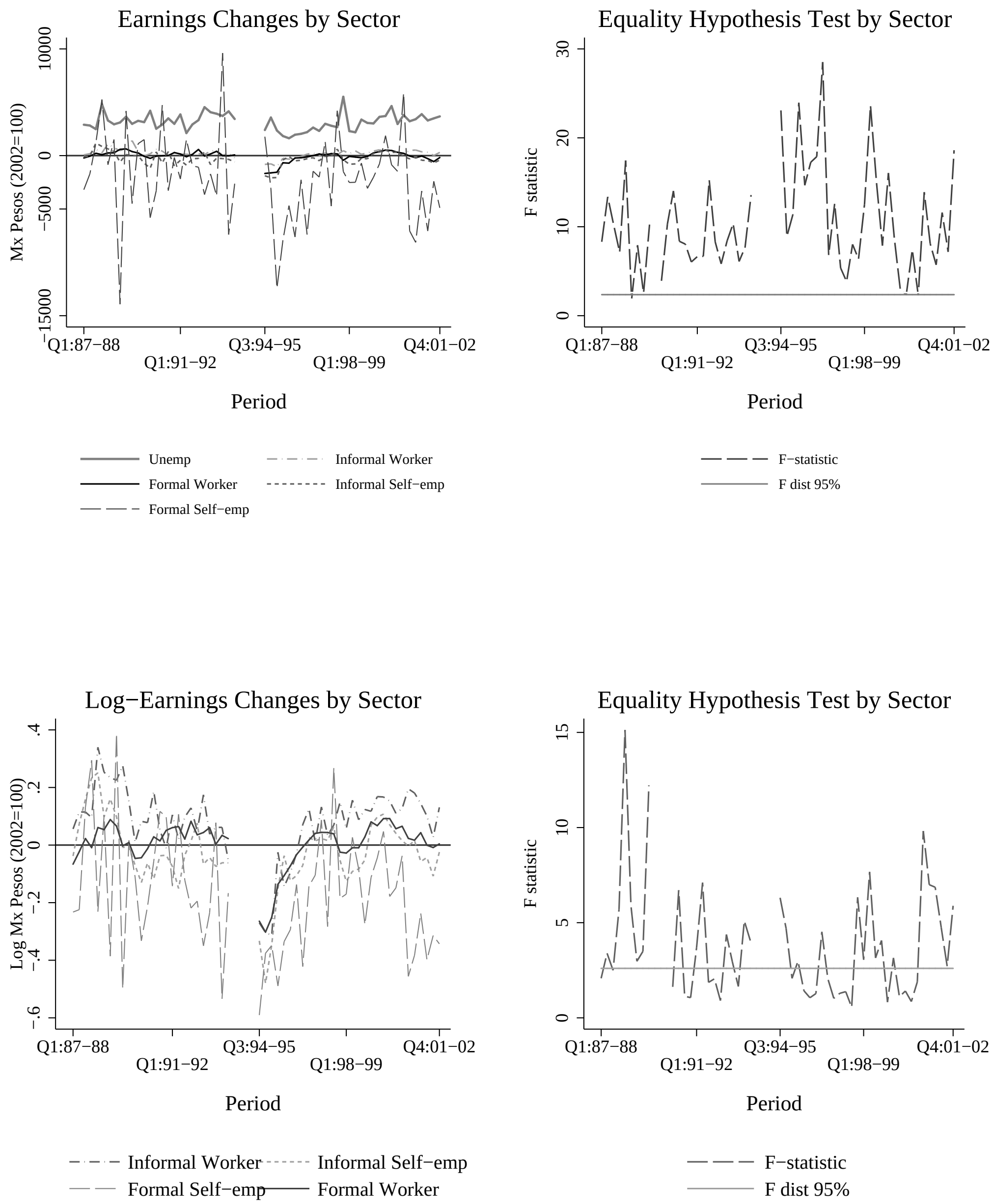


Figure 3.13: Average Mobility by Initial Sector

mobility by sector transitions. Due to the fact that certain sector transitions have very few observations on any single year, a year-by-year comparison of the mobility measures would give very unreliable estimates. For this reason several periods are pooled together according to the main trends experienced by average earnings in the economy, previously identified in Chapter 2. To recall, these periods are from the 1st quarter of 1987 to the 2nd quarter of 1993, from the 3rd quarter of 1994 to the 1st quarter of 1999, and finally from the 2nd quarter of 1999 to the 4th quarter of 2002.

The results by sector transition are shown in Tables 3.1-3.3. These tables show that becoming an informal wage worker is the worst in terms of mobility (after moving into unemployment, of course), while moving into formal self-employment is associated with the largest positive gains (and/or smallest losses). Most of the times movements into informal self-employment appear to bring larger gains (and smaller losses) than movements into formal wage work, but the differences are not large and this conclusion changes depending on whether means or medians are considered, and on the period under consideration. In particular, in the aftermath of the 1994 Peso crisis many times transitions into formal wage work are associated with less negative mobility than the ones into informal self-employment. Lacking direct measures of capital in the survey, it is hard to know whether what self-employed individuals report as earnings are the return to labor only, or whether they also include the returns to capital. The analysis of log-mobility patterns in Tables 3.1 and 3.3 leads to similar conclusions to the ones in levels.

Table 3.1: Average Earnings Mobility by Sector Transitions. All Periods.

	Levels			Logarithms		
	Mean	Std.Dev.	Median	Mean	Std.Dev.	Median
Unemployed to Unemployed	0.0	0.0	0.0			
Unemployed to Informal Worker	2249.8	1317.0	1906.4			
Unemployed to Informal Self-emp	3807.4	4030.2	2633.0			
Unemployed to Formal Self-emp	21576.8	14694.4	24339.2			
Unemployed to Formal Worker	3859.3	3407.3	2866.1			
Informal Worker to Unemployed	-2508.7	1773.9	-2144.9			
Informal Worker to Informal Worker	-52.3	1988.5	-49.1	-0.011	0.528	-0.031
Informal Worker to Informal Self-emp	763.4	2881.5	354.5	0.187	0.723	0.144
Informal Worker to Formal Self-emp	4396.1	5473.0	4063.9	0.716	0.681	0.926
Informal Worker to Formal Worker	327.9	1983.3	206.6	0.138	0.539	0.091
Informal Self-emp to Unemployed	-3810.5	4128.2	-2918.4			
Informal Self-emp to Informal Worker	-939.0	5592.3	-359.6	-0.190	0.732	-0.167
Informal Self-emp to Informal Self-emp	-135.6	6073.1	-79.0	-0.020	0.784	-0.040
Informal Self-emp to Formal Self-emp	3387.4	14717.2	1629.3	0.272	0.915	0.252
Informal Self-emp to Formal Worker	-652.3	6228.0	-30.5	0.011	0.809	0.003
Formal Self-emp to Unemployed	-10636.8	6919.7	-6791.8			
Formal Self-emp to Informal Worker	-3107.4	8210.7	-1185.8	-0.487	0.915	-0.254
Formal Self-emp to Informal Self-emp	-2405.9	15038.6	-1131.5	-0.226	0.905	-0.156
Formal Self-emp to Formal Self-emp	-518.8	22926.4	-628.2	-0.011	0.872	-0.058
Formal Self-emp to Formal Worker	-2985.7	14091.2	-2266.2	-0.279	0.808	-0.254
Formal Worker to Unemployed	-4986.6	5781.4	-3222.5			
Formal Worker to Informal Worker	-197.6	2227.5	-116.7	-0.078	0.546	-0.058
Formal Worker to Informal Self-emp	418.9	6064.0	103.3	0.003	0.733	0.029
Formal Worker to Formal Self-emp	2421.3	15522.4	1645.8	0.219	0.911	0.189
Formal Worker to Formal Worker	8.3	4443.3	1.5	0.007	0.487	0.001

Table 3.2: Average Earnings Mobility by Sector Transitions, by Period.

	Q1:87-Q2:93			Q3:94-Q1:99			Q2:99-Q4:01		
	Mean	Std.Dev.	Median	Mean	Std.Dev.	Median	Mean	Std.Dev.	Median
Unemployed to Unemployed	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unemployed to Informal Worker	3013.5	1373.3	2704.6	1883.1	1173.0	1635.3	2171.8	1147.1	1893.4
Unemployed to Informal Self-emp	4551.6	4641.1	3115.5	2951.7	2756.5	2177.1	5281.4	5474.5	3666.4
Unemployed to Formal Self-emp	17810.7	25158.8	11409.0	24520.1	10292.5	24339.2	9371.5	632.8	9683.8
Unemployed to Formal Worker	3929.1	3201.1	2963.9	3516.8	3220.4	2472.8	4578.2	4052.9	3357.5
Informal Worker to Unemployed	-2607.9	1509.5	-2352.6	-2526.7	2049.2	-2136.2	-2211.5	1064.1	-1803.3
Informal Worker to Informal Worker	-89.2	2209.6	10.0	-157.3	2165.7	-177.8	146.2	1319.3	77.5
Informal Worker to Informal Self-emp	1029.8	3526.1	487.7	512.0	2384.8	183.4	749.1	2432.4	475.0
Informal Worker to Formal Self-emp	3990.8	5207.3	5052.9	4052.9	4961.8	2176.1	6222.0	7078.1	4063.9
Informal Worker to Formal Worker	424.0	2267.1	209.2	101.5	1572.2	56.6	491.6	1996.3	392.0
Informal Self-emp to Unemployed	-4560.1	5036.5	-3497.5	-3494.0	3723.7	-2431.1	-3061.7	2333.8	-2465.2
Informal Self-emp to Informal Worker	-975.2	8264.2	-279.0	-990.4	2889.9	-477.5	-809.5	3794.6	-227.1
Informal Self-emp to Informal Self-emp	16.4	7243.4	-38.6	-378.7	5029.8	-176.0	-18.9	5327.6	30.7
Informal Self-emp to Formal Self-emp	4509.0	17794.0	1289.6	1439.3	9702.2	1811.3	3675.5	13104.2	2306.0
Informal Self-emp to Formal Worker	-872.1	6532.0	-151.0	-800.0	6607.3	-52.6	-84.2	5044.5	160.7
Formal Self-emp to Unemployed	-15498.5	6288.3	-15000.2	-8478.8	6153.2	-5750.7	-4827.6	1235.5	-4163.8
Formal Self-emp to Informal Worker	-3001.3	10291.1	-568.8	-3882.6	3179.3	-4786.1	-2181.9	5882.1	-5115.5
Formal Self-emp to Informal Self-emp	-622.4	14953.7	-850.4	-3999.2	12312.8	-1167.8	-4850.2	18287.3	-1273.5
Formal Self-emp to Formal Self-emp	-121.2	21463.7	-266.8	-2467.6	26091.5	-2389.6	1727.5	20294.0	-550.3
Formal Self-emp to Formal Worker	-2436.2	14028.1	-2055.5	-2334.3	15186.2	-2464.9	-5338.8	12159.4	-2304.0
Formal Worker to Unemployed	-5106.3	5535.8	-3304.9	-4633.3	4677.6	-3053.4	-5480.6	7784.5	-3395.9
Formal Worker to Informal Worker	-107.2	2609.9	-91.6	-386.8	1905.6	-222.9	-76.8	1956.4	-14.9
Formal Worker to Informal Self-emp	850.8	7436.6	341.0	-223.9	4390.7	-285.3	484.1	5054.0	296.1
Formal Worker to Formal Self-emp	2201.5	15584.4	1567.5	1818.3	13122.4	842.6	3334.8	17292.8	2373.2
Formal Worker to Formal Worker	156.8	4557.4	22.9	-267.0	4132.7	-108.6	127.5	4638.7	145.4

Table 3.3: Average Log-Earnings Mobility by Sector Transitions, by Period.

	Q1:87-Q2:93			Q3:94-Q1:99			Q2:99-Q4:01		
	Mean	Std.Dev.	Median	Mean	Std.Dev.	Median	Mean	Std.Dev.	Median
Informal Worker to Informal Worker	0.006	0.572	0.005	-0.078	0.517	-0.107	0.070	0.476	0.044
Informal Worker to Informal Self-emp	0.236	0.787	0.207	0.122	0.696	0.102	0.215	0.651	0.161
Informal Worker to Formal Self-emp	0.678	0.755	1.007	0.704	0.612	1.141	0.871	0.421	0.839
Informal Worker to Formal Worker	0.136	0.537	0.093	0.086	0.557	0.035	0.213	0.507	0.176
Informal Self-emp to Informal Worker	-0.123	0.732	-0.098	-0.272	0.746	-0.245	-0.157	0.699	-0.087
Informal Self-emp to Informal Self-emp	0.012	0.814	-0.015	-0.083	0.767	-0.093	0.020	0.750	0.014
Informal Self-emp to Formal Self-emp	0.274	0.946	0.195	0.238	0.965	0.283	0.313	0.785	0.348
Informal Self-emp to Formal Worker	-0.021	0.850	-0.039	-0.004	0.832	-0.014	0.085	0.696	0.069
Formal Self-emp to Informal Worker	-0.252	0.929	-0.231	-0.791	0.643	-1.229	-0.797	1.094	-1.602
Formal Self-emp to Informal Self-emp	-0.137	0.903	-0.156	-0.328	0.810	-0.150	-0.313	1.022	-0.205
Formal Self-emp to Formal Self-emp	0.010	0.857	-0.028	-0.054	0.938	-0.160	0.012	0.791	-0.054
Formal Self-emp to Formal Worker	-0.218	0.786	-0.232	-0.266	0.904	-0.290	-0.450	0.664	-0.266
Formal Worker to Informal Worker	-0.048	0.553	-0.050	-0.156	0.542	-0.129	-0.017	0.530	-0.008
Formal Worker to Informal Self-emp	0.067	0.741	0.090	-0.119	0.736	-0.135	0.048	0.693	0.097
Formal Worker to Formal Self-emp	0.201	0.887	0.267	0.105	0.790	0.077	0.345	1.030	0.180
Formal Worker to Formal Worker	0.022	0.487	0.007	-0.039	0.480	-0.039	0.044	0.491	0.045

3.5.3 Mobility as an Equalizer of Longer-Term Earnings

This section presents the results of the calculations of the P-index of Mobility as an equalizer of longer-term earnings for the whole sample and subgroups of the population.

In Figure 3.14 the P-index as described by equation (3.3) is plotted. The index is calculated according to two inequality measures, namely the Gini index and the Generalized Entropy index with coefficient 2 (GE(2)), which equals half the squared coefficient of variation. Although the two P-indices that arise from these calculations (denoted from now on P_{Gini} and $P_{GE(2)}$) have very different scales, they both indicate that during the years under study mobility equalized earnings. These two indices show a roughly constant pattern, but around different levels (the $P_{GE(2)}$ shows higher equalization, but higher fluctuations too). Only during a couple of periods in the sample the conclusions of the indices differ. During the late eighties the $P_{GE(2)}$ shows that mobility was disequalizing, while the P_{Gini} shows the opposite. Since the $P_{GE(2)}$ coefficient has the advantage of being decomposable into within and between-group components, the analysis in this section will continue analyzing only this index for the different subgroups of the population.

Figure 3.15 plots the P-index by initial quintile groups. The graph contains both the within and between-group components, as well as their shares (i.e., $\kappa_w P_w$ and $\kappa_b P_b$). The conclusion stemming from this graph is that in general earnings mobility helped in reducing between and within-quintile longer-term earnings inequality. However, whenever the overall mobility process is disequalizing this effect comes from disequalizations in earnings within-quintiles. In the next chapter, it will be shown that this equalization of longer-term earnings between groups occurs because the top and bottom quintiles of earnings profiles approach over time.

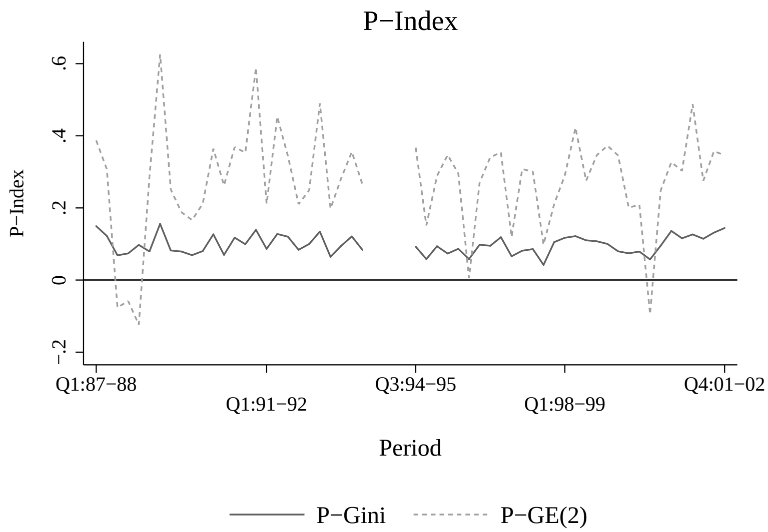


Figure 3.14: P-index

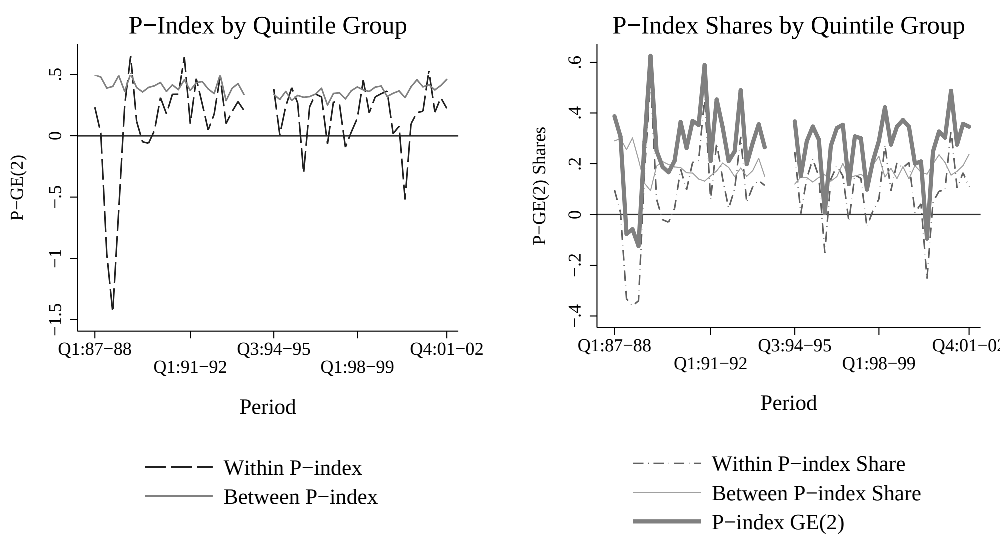


Figure 3.15: P-index by Initial Quintile Group

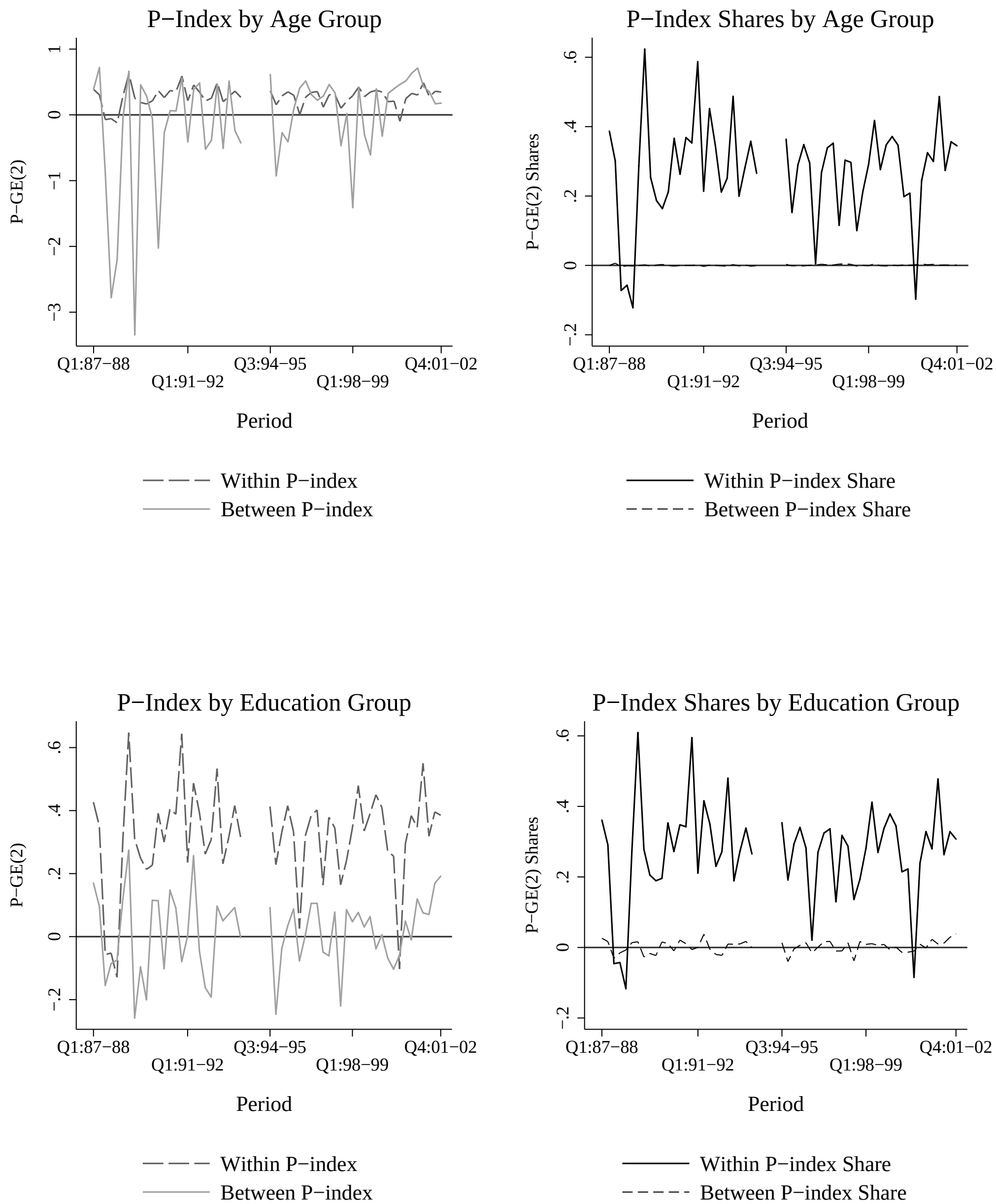


Figure 3.16: P-index by Age and Education Groups

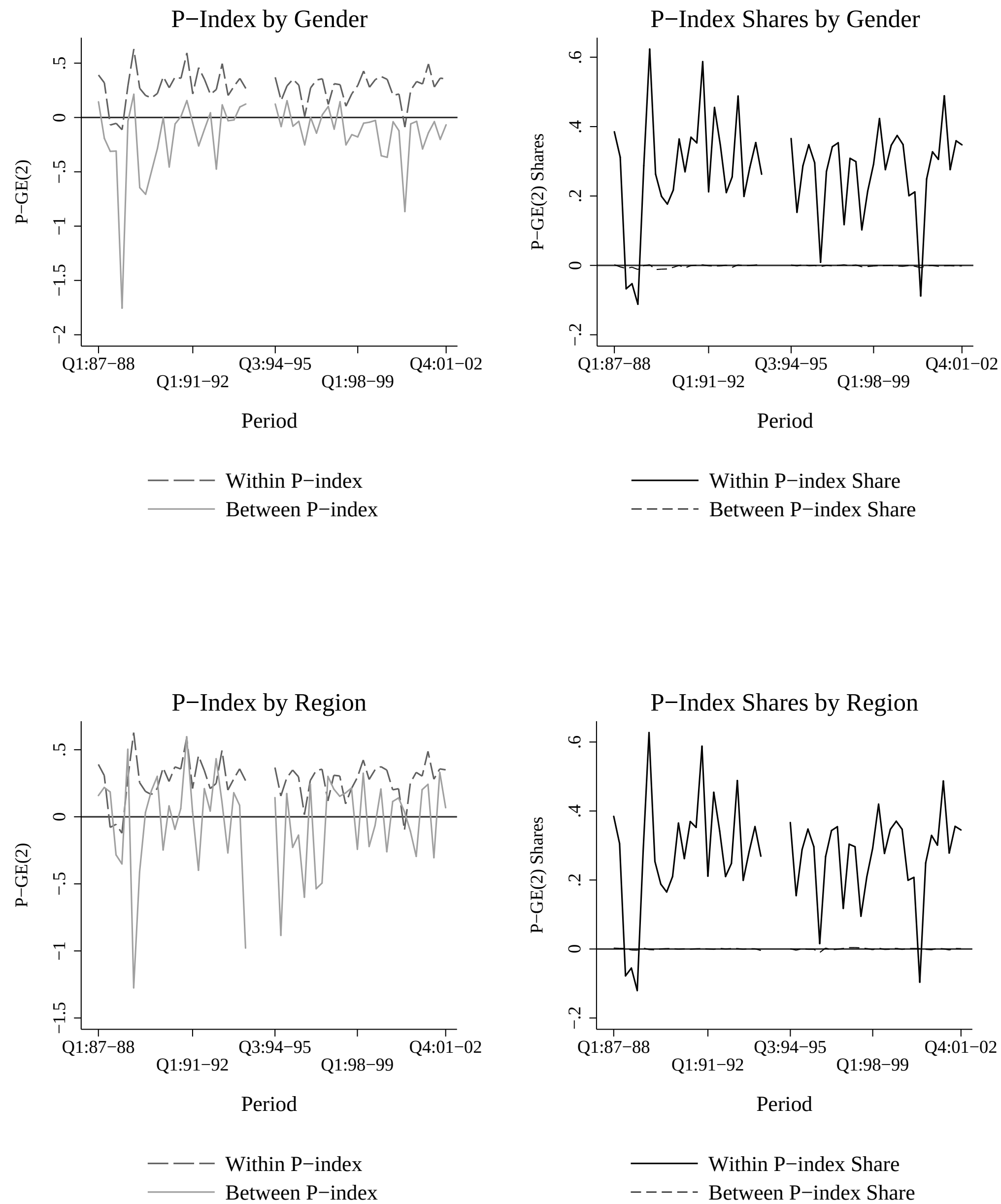


Figure 3.17: P-index by Gender and Region

The analysis of the P-index by age and education groups shows that mobility equalized longer-term earnings *within* groups; while its effect in the the between-group inequality is sometimes equalizing and sometimes disequalizing. However, when weighted by their respective weights, the equalization that took place within-groups has a dominating effect, giving the overall pattern of equalization previously noted for the entire population.

The analysis of Mobility as an equalizer of longer-term earnings by gender presents some interesting results. In Figure 3.17 it can be seen that while it is still true that earnings mobility helped equalize the longer-term earnings within each gender, its effect led to stronger earnings inequality between men and women over time. Again, once the weights of each effect are taken into account, the equalization that took place within-genders dominates the disequalization that occurred between the earnings of men and women.

Finally, in the case of earnings mobility by region, the conclusions are similar to the education and age group cases. Earnings mobility equalizes longer-term earnings within regions, while sometimes equalizes and other disequalizes longer-term earnings between regions, the former effect being the dominating one.

Unlike the concept of directional mobility, the analysis of mobility as an equalizer of longer-term earnings was not performed by initial sectors, because there were not enough observations in the all categories as to obtain reliable estimates of the within-group inequality indices.

3.6 Conclusions

This chapter studied two groups of aggregate mobility questions, namely:

1) "What are the average earnings gains and losses in the economy?" and "Are these earnings mobility patterns the same for different groups of the population?"

2) "Does mobility equalize earnings over time?", "Does mobility equalize earnings within groups over time?" and "Does mobility equalize earnings between groups over time?"

The answer to the first question is that on average earnings mobility fluctuated around zero, with the exception of the late eighties and the early 2000s, when individuals experienced upward mobolity, and the years following the 1994 Peso crisis that were associated with large earnings losses. These findings occur both in levels of earnings, as well as in logarithms.

The analysis by subgroups of the population shows that mobility differs only by initial earnings quintile groups and by sector. In both cases, the group with higher initial earnings (the fifth-quintile and the formal self-employed, respectively) experienced the largest losses, both in absolute and proportional terms. Similarly, the group that had the smallest initial earnings (the first-quintile, and both the unemployed and the informal wage workers) experienced the largest gains.

Age, education, gender and region groups, did not present major differences among themselves in Directional mobility over time. In general, more educated individuals experienced larger fluctuations in their earnings, and together with males they experienced larger losses after the Peso crisis. However, these differences disappear once their initial earnings level is accounted for, i.e. their higher mobility is proportional to their higher initial earnings level.

The answer to the second set of questions is that mobility helped in equalizing earnings for most of the periods studied. Only during a couple of periods in the late eighties mobility acted to disequalize earnings over time. It is interesting to note that it was during this same period that earnings were growing fast.

Another finding was that in general mobility helped equalize within-group longer-term earnings, while its role regarding between-group inequality was sometimes equalizing and sometimes disequalizing. However, in most of the cases the equalization at the the within-group level dominated the other one. The only exception to this finding occurred when initial earnings quintile groups were analyzed. In this case, earnings mobility helped equalize between-group longer-term earnings. One interesting finding was that earnings mobility almost always increased the inequality between the longer-term earnings of men and women, but the equalization that took place within gender groups dominated at the end.

A natural direction in which to move from here is to extend the analysis of Directional mobility by controlling for more than one factor at a time. This is done in the next chapter where the determinants of earnings mobility are analyzed. Before that though, a closer look will be given to the relationship between initial advantage and earnings mobility.

Chapter 4

Initial Earnings and the Determinants of Earnings Mobility

4.1 Introduction

This chapter studies the impact of initial earnings on mobility, as well as the effect of other socioeconomic variables on earnings changes. In particular, the questions addressed are: "Are the most advantaged individuals gaining more (losing less) in terms of earnings changes?", "What is the *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility?", and finally "How do these factors affect the impact of initial earnings on mobility?".

The first question is concerned with whether the mobility process benefits(hurts) the rich more(less) than the poor, or is it that this process benefits more the low-earners, allowing them to catch-up as time goes by? This question is closely related to the study of poverty traps and cumulative advantage. The main difference between those studies and this chapter is that they are usually concerned with dynamic processes in the long-run, while this chapter, due to data limitations, focuses on mobility in the short-run. Studying mobility in the short-term allows identifying the differential impact of macroeconomic shocks on the mobility of individuals at different points of the earnings distribution.

With respect to the second set of questions, i.e., the study of the determinants of earnings dynamics, this chapter tries to specify which variables, in addition to initial earnings, explain earnings mobility. In particular, it identifies the impact of factors like gender, education, age, sector of employment (informal vs. formal), and geographical region on earnings changes.

Two methodological issues that are of particular concern in the mobility literature are the potential negative effects of measurement error in the earnings variable and the attrition of individuals from the panel. Both topics are considered in this chapter and the robustness of the results to these problems is explored.

Section 4.2 summarizes previous findings in the literature and presents the contribution of this chapter. The methodology followed is explained in section 4.3, and results are presented in section 4.4. Section 4.5 concludes.

4.2 Previous Research and Contribution

The literature on the relationship between mobility and initial earnings focuses on two different questions. The first one is "What is the relationship between earnings changes and initial earnings?", and the second "What is the relation between earnings changes and initial earnings, after one has controlled for the effects of individual characteristics like age, education, gender, etc.?".

These two questions are very different in nature. The unconditional question deals with the common concern of whether "the richer are getting richer (and the poor poorer)", while the conditional one is concerned with the determinants of mobility and the existence of state dependence in the conditional earnings dynamics.[1]

For developed countries, the study of earnings dynamics has been pursued at a great level of detail. In the US alone, earnings mobility studies have addressed issues like the role of on-the-job-training on earnings (Hause, 1977), poverty dy-

[1]Later in this chapter it will be seen that this question implies studying whether an individual converges to his own permanent level of earnings.

namics (Lillard and Willis, 1978), wage dynamics and job turnover (Lillard, 1999), and the covariance structure of earnings *per se* (Lillard and Weiss, 1979; MaCurdy, 1982; Abowd and Card, 1989). The high quality of the data in these countries has allowed mobility researchers to even explore the dynamics of income variance (Meghir and Pistaferri, 2001), and the effects of measurement error on the estimated earnings mobility by means of validation data (Pischke, 1995).

For the case of developing countries the panorama is less positive. Most of the panel data for these countries have few observations over time, and hence many mobility studies are performed using two temporal observations per unit of analysis (see for instance Grootaert *et al.*, 1997; Fields *et al.*, 2003a,b).[2] In addition, the lack of validation studies in these countries makes hard to assess the extent of measurement error on the earnings variable, and its potential impact on the estimated mobility parameters. Finally, another issue that makes the mobility research harder to perform in these countries is the existence of high levels of attrition in the surveys collected. In spite of these difficulties, research on earnings mobility has continued to grow in the developing world.

For the Mexican economy two recent studies have appeared dealing with issues very close in spirit to the present chapter. The papers by Antman and Mckenzie (2005a) and Antman and Mckenzie (2005b) study earnings dynamics with the same data over similar periods of time. Since the ENEU consists of 1-year panels, and the authors are interested in studying mobility in the long-run, they create pseudo-panels in which specific age-education cohort groups are tracked over long periods of time. This method has advantages in extending the temporal coverage for mobility studies over many years, and it potentially helps to mitigate the problems

[2]Needless to say, this clearly limits the type of dynamic structures that can be estimated.

of measurement error and attrition bias. However, this methodology makes strong assumptions that are problematic in practice. First, by tracking the mobility of a cohort they miss the study of any intra-cohort mobility that might take place over time. Second, one cannot be sure that the mobility experienced by a cohort group actually represents the true mobility experienced by a given group of individuals. Issues like migration, deaths and household dissolution and creation might lead to incorrect inferences when this method is applied. As rightly pointed out by Deaton when discussing this methodology (otherwise strongly advocated by him) "(...) time series of cross sections can tell us about average earnings for the cohort over time, and it can tell us about inequality of earnings within the cohort and how it is changing over time, but it cannot tell us how long individuals are poor, or whether the people who are rich now were rich or poor at some earlier date" (Deaton, 1997, p.120).

In Antman and Mckenzie (2005a) the authors focus on mobility in household labor income. The authors are interested in studying whether there is unconditional convergence between the earnings of rich and poor households (what the authors call absolute convergence), and whether there is conditional convergence of the household's earnings to its own average level. Since they work with cohort average incomes, it is important to keep in mind that all the mobility results here reviewed correspond to such averages. The authors find very little absolute convergence between rich and poor households, i.e., in general households keep their income levels over time. However, there is rapid conditional convergence, and it increases as time goes by.

The authors also compare pseudo-panel quarterly mobility estimates to the ones stemming from true panels (following individuals instead of cohorts) and they find that their pseudo-panel results are surprisingly similar to the ones obtained through instrumental variable estimations (attempting to correct for measurement error) in the true panel. These estimates show much slower convergence than the ones obtained through Ordinary Least Squares.

According to the authors, analyzing mobility over cohort averages gives them the advantage of solving the problems of measurement error and attrition bias commonly encountered in this type of studies. Regarding the measurement error problem, although it seems plausible that averaging the incomes of several households in a given cohort will tend to diminish the household idiosyncratic measurement error, this may not solve the overall problem if the households in a given cohort systematically misreport their earnings, e.g., if households with highly educated middle-aged heads underreport their income. As a solution to the attrition problem, the authors use the first interview for each household (when there is no attrition) in the construction of their pseudo-panels. Their overall finding is that there is no substantial difference in their convergence estimates for subsamples of attritors and non-attritors.

In a companion paper (Antman and Mckenzie, 2005b) the authors use the same pseudo-panels to study whether there are poverty traps in Mexico. More specifically they study the possibility of nonlinearities in household labor-income dynamics. Their finding is that there are no poverty traps for Mexican urban households. One limitation of this study is that it bases its conclusions on the analysis of quarterly dynamics of income. Hence, the authors try to capture poverty traps under a very stringent definition of what these traps are.

Other studies analyzing the determinants of earnings mobility in Mexico are Maloney and Cunningham (2000), Maloney *et al.* (2004), and World Bank (2004).

The main aim of this literature is studying vulnerability in Mexico. In particular, they ask which subgroups of the population are more "vulnerable" to income falls.

Their analysis focuses mainly on studying what happens at the bottom of he conditional earnings mobility distribution, where the conditioning factors are a set of socioeconomic variables. The authors of these studies consider that the group at the bottom of the conditional earnings distribution suffers "disproportionate 'catastrophic' shocks" (i.e., they are more "vulnerable") and these shocks would not be captured by standard regression methods.

The periods covered by Maloney and Cunningham (2000) and Maloney *et al.* (2004) include before, during and after the 1994 Peso crisis, as well as 2000-2002. The data set used by these studies is again the ENEU. Among their main findings are that, holding everything else constant, the least educated and poor suffered slightly less in terms of earnings changes during the 1994 Peso crisis, but at the cost of having to put other members of the household in the labor market. They also find that if higher weights are attached to the income changes of poorer households, the households with a less educated head present large losses, something interpreted by the authors as higher vulnerability. Finally, they find that the structure of the determinants of earnings changes is quite stable regardless of whether the economy is in recession or not. The only main difference observed is that, holding everything else constant, during recessions the more educated households experience larger earnings losses.

It is important to stress that none of the conditional mobility estimations in these papers included the initial income level as an explanatory variable. Nevertheless, some evidence is provided for the relationship between household income change and a proxy for permanent income. The relationship they find between these two variables is negative and stronger during the 1994-95 recession (when compared to the one of the recovery period that followed afterwards).

One interesting analysis conducted in World Bank (2004) is the inclusion of rural households. This study incorporates results based on a recently created rural panel survey that complements the ENEU to form the new National Survey of Quarterly Employment (ENET). The period of analysis goes from 2000 to 2002. The results obtained with the ENET are compared to the ones from another rural panel generated to evaluate the PROGRESA poverty alleviation program. This last panel covers the 1998-2000 period and, in contrast with the ENET, it contains information on consumption of the households. While the authors obtain similar results when comparing the urban and rural sub-samples of the ENET, they reach very different conclusions when analyzing consumption changes in the PROGRESA panel. In particular, they find considerable consumption smoothing on the part of the households. More importantly, they find that less educated households and household headed by older workers fare worse in terms of consumption changes. Since these categories are proxies for a permanent disadvantage, it seems that more disadvantaged households fared worse in terms of consumption mobility. These results contradict the findings for the rural part of the ENET, and stress the fact that income (and its change) cannot fully appraise the welfare of a group of individuals. Another important conclusion coming out of this comparison is that what happens in rural Mexico in terms of economic mobility, might be very different from what happens in urban areas. In particular, there is a generalized perception that the rural areas have fared worse in terms of mobility in Mexico. The findings reported in World Bank (2004) provide evidence supporting this perception. Since this chapter focuses on urban areas only, it is crucial to keep in mind that the findings here presented may not generalize to the whole country.

In the light of these previous studies, the contribution of the present chapter to the previous mobility literature is to focus in the short-run earnings dynamics experienced by individuals in urban Mexico, with an emphasis on the role played by initial advantage on mobility. This chapter also provides further evidence on the role played by socioeconomic factors determining mobility, and interprets the results within the framework of a structural model of earnings. The results obtained are analyzed over a long period of time, with varying macroeconomic conditions. Finally, the robustness of the findings to different measures of initial advantage, to measurement error and to attrition bias are tested. Some results similar to the ones here reported also appear in Fields *et al.* (2005).

4.3 Methodology

4.3.1 Unconditional Mobility

This section introduces the methodology used to analyze the relationship between mobility and initial advantage. Denote by y_{it} the earnings of individual i at period t, and its change by Δy_{it}, then in order to answer the question: "Are the most advantaged individuals gaining more (losing less) in terms of earnings changes?", one natural place to start is by estimating the expected earnings changes given initial earnings, i.e., $E(\Delta y_{it}|y_{it-1})$. The simplest assumption to make about this conditional expectation is that it is linear, i.e.,

$$\Delta y_{it} = \beta_0 + \beta_1 y_{it-1} + u_{it} \tag{4.1}$$

Then the answer to the previous question will depend on the sign of the β_1 parameter. More specifically, there will be divergence in earnings if $\beta_1 > 0$. On the other hand there will be convergence if $\beta_1 < 0$, and the earnings changes patterns will be parallel if $\beta_1 = 0$. In other words, a positive β_1 means that the individuals at the top of the initial earnings distribution have more positive (or less negative) earnings changes, i.e., the rich got richer. On the contrary, if β_1 is negative then there is convergence between the individuals at the bottom and the ones at the top of such distribution, i.e., the least advantaged gain the most (lose the less). Finally, if β_1 equals zero then earnings mobility is not affected by initial earnings, and mobility depends only on the constant β_0 and the random factors captured by u_{it}. Since these random factors average to zero, a $\beta_1 = 0$ means that on average everybody experiences the same mobility β_0.

The relationship stated in equation (4.1) can be estimated by Least Squares (LS) for earnings both in levels and logarithms. The estimation in levels gives a measure of the convergence in pesos, while the logarithmic specification estimates the amount of log-convergence, which gives a larger weight to the mobility of poorer individuals and approximates the proportional mobility by level of initial earnings.

Since there are many overlapping short-run panels over which to estimate this relationship, there will be several LS estimates of the β_1 parameter, one for each period (i.e., there will be many β_{1t}'s). With these many β_{1t}'s it is possible to track the evolution of convergence over time, and across varying macroeconomic conditions. Since outliers might create a problem in certain years, a median regression is also estimated and the results are compared to the ones of the LS analysis.[3]

The interpretation of the previous β_1 parameters is an issue that deserves further discussion. Even if it were observed that earnings converge (i.e., $\beta_1 < 0$) it is

[3]As mentioned in the Chapter 2 all the parameters reported are weighted estimates using sampling weights, and the standard errors account for the survey design in the ENEU.

not evident what meaning should be attached to this finding. A negative β_1 could be the product of reversion to the mean resulting from adjustments in earnings to a temporary shock. For instance, it is possible that individuals who reported having low (high) earnings in the base period were temporarily unlucky (lucky) and that the positive (negative) mobility observed for them is just an adjustment back to their permanent earnings level.[4] A negative β_1 could also mean that individuals at the bottom are truly faring better by experiencing gains that will continue in the future.

As previously mentioned, issues of mobility in the long-run cannot be analyzed with the data at hand. However, measures of permanent advantage can be generated and used to analyze their impact on mobility. In particular, a regression similar to (4.1) can be estimated using a proxy of permanent advantage as the independent variable. In this chapter, this predicted permanent advantage measure $\hat{y}_{it-1}$ is formed in two ways: first, by averaging the earnings of the individual using all the quarters of information available (instead of using earnings from the first interview only), and second, by regressing initial earnings on variables that would be good predictors of the permanent advantage of an individual. These predictors include human capital variables like age, education, gender, and wealth proxies like cluster average earnings, and dwelling characteristics. The prediction exercises are done by Least Squares and by median regression, one for each period, and new β_{1t}^{P}'s are estimated using each one of these methods.

[4] Further empirical evidence of whether this is happening will be provided below.

4.3.2 Conditional Mobility and the Socioeconomic Determinants of Mobility

This section presents the methods used to analyze the second set of questions: "What is the *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility?" and "How do these factors affect the impact of initial earnings on mobility?".

Broadly speaking, the socioeconomic characteristics of an individual can be grouped in time-invariant characteristics Z_i and time-variant characteristics X_{it-1} and X_{it}. The time-invariant characteristics include gender, age (linear and squared), education (linear and squared), and regional dichotomous variables. The time-variant variables refer to sector of employment, meaning formal wage work, formal self-employment, informal wage work, informal self-employment, and unemployment.

In order to be able to interpret the results of the conditional mobility estimations within a structural framework, it is useful to start with specifying an earnings equation and from there derive a mobility equation. A natural starting point is to allow earnings at time t to be affected by all the factors listed under Z_i and X_{it}. That means that earnings are determined by age, gender, education level, region, and sector of employment.

Whether this constitutes truly exogenous "determination" or not, is a matter of debate. Variables like sector of employment and region are potentially endogenous to the earnings determination process, since an individual could choose which sector to work in, or which region to migrate to depending on his earnings mobility. Keeping this caveat in mind, this section will proceed by treating these variables as pre-determined, and will ignore the potential complications that arise due to these issues.[5] Finally, since no attempt will be made to correct for the potential self-selection of individuals into the labor force, all the results should be considered to apply only for the subpopulation of individuals participating in the labor force both in the initial and the final periods.

The basic specification of the earnings equation is[6]

$$y_{it} = Z_i\gamma_t + X_{it}\kappa_t + \varepsilon_{it} \tag{4.2}$$

where the error term ε_{it} is independent of Z_i and X_{it} and autocorrelated, i.e.,

$$\varepsilon_{it} = \rho_t\varepsilon_{it-1} + \eta_{it}$$

$$\eta_{it} \perp X_{it}, Z_i \qquad\qquad \eta_{it} \sim (0, \sigma_\eta^2) \qquad \forall i, t$$

This equation states that earnings at time t depend on the time-invariant characteristics Z, time-variant characteristics X at time t, and an error term ε that captures shocks to earnings. The effects of past values of the time-variant characteristics and of the shocks are assumed to enter only through the current values of these factors. This equation provides a rationale for why initial earnings would affect earnings changes, even after conditioning for socioeconomic characteristics. In particular, the AR(1) structure assumed for the ε term ensures that y_{it-1} has an impact on earnings mobility. To see why note that the earnings changes implied by (4.2) are

$$\Delta y_{it} = Z_i(\Delta\gamma_t) + (\Delta X_{it})\kappa_t + X_{it-1}(\Delta\kappa_t) + ((\rho_t - 1)\varepsilon_{it-1} + \eta_{it})$$

hence substituting $\varepsilon_{it-1} = y_{it-1} - Z_i\gamma_{t-1} - X_{it-1}\kappa_{t-1}$ into this expression leads to

$$\Delta y_{it} = (\rho_t - 1)y_{it-1} + Z_i\tilde{\gamma}_t + (\Delta X_{it})\kappa_t + X_{it-1}\tilde{\kappa}_t + \eta_{it} \tag{4.3}$$

[5]Chapter 5 in this dissertation studies more closely the issue of sector selection, in particular whether individuals are free to choose among sectors and the implications for earnings mobility.

[6]A similar model was used in Fields *et al.* (2003a)

where $\tilde{\gamma}_t = \gamma_t - \rho_t\gamma_{t-1}$ and $\tilde{\kappa}_t = \kappa_t - \rho_t\kappa_{t-1}$. Therefore, under this model, the effect of initial earnings y_{it-1} on earnings mobility after conditioning a on a set of socioeconomic variables comes from the autocorrelation of the unobserved error component.[7]

In the present application of the model described by eq. (4.3), the only time-varying variables will be dichotomous variables indicating the sector of employment, as a result, a slightly modified version of this equation is estimated. In particular, denote by $st(l,m)$ a dichotomous variable that takes value 1 if the individual transited from sector l to sector m, and zero otherwise, and let $\pi_{lm} = \kappa_t(m) - \rho_t\kappa_{t-1}(l)$, where $\kappa_t(j)$ is the *j-th* element of the vector κ_t, i.e., is the parameter for the sector j in the earnings equation (4.2). Under this notation, the term $(\Delta X_{it})\kappa_t + X_{it-1}\tilde{\kappa}_t$ equals $\sum_l \sum_m st(l,m)\pi_{lm}$, hence the conditional mobility equation (4.3) can be rewritten as

$$\Delta y_{it} = (\rho_t - 1)y_{it-1} + Z_i\tilde{\gamma}_t + \sum_l \sum_m st(l,m)\pi_{lm} + \eta_{it}. \tag{4.4}$$

This is the equation that will be estimated.

This model subsumes the partial adjustment model as a particular case. If the steady state earnings of an individual are defined as

$$y_i^s = Z_i\gamma + X_i\kappa$$

and if $X_{it} = X_i, \kappa_t = \kappa, \gamma_t = \gamma, \rho_t = \rho$, then equation (4.4) can be rewritten as

$$\Delta y_{it} = (1-\rho)(y_i^s - y_{it-1}) + \eta_{it}$$

[7]A richer version of the model expressed by (4.2) and (4.3) would allow for the presence of individual unobserved time-invariant effects, δ_i. Since the present dissertation focuses on yearly mobility, of which only one observation per individual is available, it will not be possible to separately identify the aforementioned effects. For this reason, this more complicated structure is not pursued here. Nevertheless, the reader should keep in mind this limitation when interpreting the parameters of variables under X and Z estimated from eq. (4.4).

where the parameter ρ is adjustment coefficient of earnings to its steady state. In particular, if the variables grouped under X and Z constitute the determinants of permanent earnings, the influence of y_{it-1} on earnings mobility comes from the adjustment of earnings to its permanent level.

The issue of which variables constitute determinants of permanent earnings deserves careful consideration. While it is less controversial that time-invariant variables like age, education and gender are determinants of long-term advantage (and hence of permanent earnings), characteristics like sector of employment and region do not necessarily determine permanent earnings. Whether they do or not depends on how easy it is for individuals to move geographically and across sectors.[8]

Equation (4.4) is estimated by (robust) LS and median regression for changes in earnings and log-earnings under several specifications. The specifications include: (1) Human Capital variables (age and education), plus gender and regional dummies as the only regressors, (2) the previous variables plus sector transitions dummies, and (3) the variables in specification (2) allowing ρ_t to vary for different groups of the population. The groups considered are gender, age, education, initial sector of employment and region. Specification (3) tests whether there are different rates of convergence for each of these groups.

Since the methods presented in this section involve the comparison of large amounts of results, the presentation of such results will be done by graphing the coefficient of y_{it-1}, i.e., $(\rho_t - 1)$ for the various specifications, and the full set of regression results will be presented only for the data grouped under 3 pooled periods. The pooled periods are Q1:87-Q2:93, Q3:94-Q1:99, and Q2:99-Q4:02.

[8]The issue of geographic mobility is not dealt with here, since the data does not allow tracking individuals that migrate.

4.3.3 Robustness checks

Two issues that concern many researchers studying mobility are the presence of measurement error in the earnings variable and attrition bias. Errors or misreports of earnings can lead to serious biases in the estimation of the coefficients in equations (4.1) and (4.4). It could even be the case that initial earnings have no effect on mobility, and still a relationship is found due to the correlations of the measurement error terms. On the other hand, the existence of attrition (and non-reporting) leads to problems in identifying the conditional expectations of interest, since the dependent variable will not be observed for a fraction of the population. If both problems present themselves together then there could be a serious threat to the quality of the estimates. This section presents methods to assess the possible impacts of these two problems, one at a time.

4.3.4 Measurement Error

Until recently, it was usually assumed that measurement error of economic variables was always of the classical variety, i.e., the measurement error was assumed to be an iid term, uncorrelated with the true value of the variable of interest, with any other variable in the model, with the error term in the equation of interest, and with any other measurement error in any other variable. Although this model is analytically convenient, enough evidence has accumulated over the past decade showing that this assumption does not hold in general for the earnings variable used in labor studies (see Bound and Krueger, 1991; Bound *et al.*, 1994; Pischke, 1995; Bound *et al.*, 2001). Individuals tend to report what they "usually" earn, and not necessarily the exact earnings they had in a specific period. Also, reach individuals might underreport their earnings out of fear that the survey will be used for tax purposes.[9] Similarly, individuals at the bottom of the earnings distribution might overstate their incomes out of embarrassment. Unfortunately, for the case of Mexico there are no validation studies that allow testing the nature of this potential problem.

Given this data limitation, the approach taken here is to follow the measurement error model proposed in Bound *et al.* (1994) and extended by Pischke (1995), in order to show some implications of this model for the mobility estimations performed in this chapter.[10] The main caveat of proceeding this way is that the measurement error evidence on which this section relies, comes from a validation study performed on a single Detroit firm in the mid-eighties.[11] The earnings measure in that study is annual earnings coming both from employer records and the answers to a PSID questionnaire applied to a sample of workers in that firm.

Clearly, using a validation study for a single US firm is far from satisfactory (although a similar model applies to the more representative CPS survey when compared to Social Security records). Also, in the case of Mexico the ENEU reports monthly earnings, instead of yearly earnings. This earnings variable is constructed by allowing the respondent to choose a preferred time framework (day, week, month, etc.) and to report their earnings during that period. After that, the interviewer performs whatever conversion is necessary to transform that report into monthly earnings. Although this scheme reduces the error due to bad recall of true earnings, it makes the PSID model less applicable to the Mexican case.

[9]In the case of Mexico, it seems more plausible that such individuals would underreport their earnings out of fear of being kidnaped.

[10]A similar route was adopted before by Fields *et al.* (2003a) following a variant of the model proposed by Bound *et al.* (1994).

[11]The validation study is the PSID Validation Study, for a description see Duncan and Hill (1985) and Duncan and Mathiowetz (1985).

The measurement error model proposed in Bound *et al.* (1994) is one where the measurement error has mean zero and is "mean reverting", i.e., it is negatively correlated with the true value of earnings. Later on, Pischke (1995) studied more carefully the same validation survey and the relationship between measurement error and earnings dynamics. He proposed a slightly different version of the Bound *et al.* (1994) measurement error model, in which the mean reverting measurement error term applied only to the transitory component of earnings, i.e., when asked about their earnings for the preceding year individuals tended to report their usual earnings.

Based on the previous insights, the following measurement error model is used in this dissertation. Let y_{it} be measured earnings, y^*_{it} be the true value of current earnings and μ_{it} be the measurement error. Then

$$y_{it} = y^*_{it} + \mu_{it}. \tag{4.5}$$

True earnings are assumed to be the sum of two orthogonal components, permanent earnings, y^P_{it}, and transitory earnings, which in order to follow the notation established in (4.2) will be denoted by ε_{it},[12]

$$y^*_{it} = y^P_{it} + \varepsilon_{it} \qquad\qquad y^P_{it} \perp \varepsilon_{it} \tag{4.6}$$

Furthermore, the measurement error is assumed to be linearly related to true earnings according to the following equation[13]

$$\mu_{it} = \alpha(y^*_{it} - y^P_{it}) + \zeta_{it} \qquad\qquad \alpha < 0 \tag{4.7}$$

[12]Note that the term denoting permanent earnings y^P_{it} is not necessarily constant over time, making it consistent with the interpretation given under eqn. (4.2).

[13]Both in Bound *et al.* (1994) and Pischke (1995) the measurement error model is derived for log-earnings rather than earnings in levels; however, Pischke (1995) notes that the same structure also applies to the earnings variable in levels.

where the term ζ_{it} is the idiosyncratic component of measurement error which has mean zero, finite variance σ^2_ζ, and it is uncorrelated with true earnings, but it can be autocorrelated. In particular, ζ_{it} is assumed to follow an AR(1) process, i.e.,

$$\zeta_{it} = \theta\zeta_{it-1} + \omega_{it} \qquad\qquad \omega_{it} \sim iid(0, \sigma^2_\omega) \tag{4.8}$$

If the measurement error follows the previous structure, then it can be shown that the OLS estimate of β_1 in eq. (4.1) will be given by

$$\hat{\beta}_1 = \beta_1 \frac{V(y^*_{it-1})}{V(y_{it-1})} + (\rho - 1)\frac{\alpha(2+\alpha)V(\varepsilon_{it-1})}{V(y_{it-1})} + (\theta - 1)\frac{V(\zeta_{it-1})}{V(y_{it-1})} \tag{4.9}$$

as shown in the appendix of this chapter.[14]

Expression (4.9) can be used to make simple simulations by assuming possible values for ρ, θ and α. The simulations performed will aim at estimating what size of measurement error would lead to the erroneous conclusion that there is convergence in earnings, when in fact the true β_1 is zero.

Another result that follows from this assumed structure of measurement error, is that the estimated coefficient when regressing Δy_{it} on predicted permanent initial advantage $\hat{y}_{it-1}$ will not be biased. A formal derivation is also included in the appendix.[15]

Finally, turning to the implications of measurement error on the conditional mobility estimations, a substitution of equations (4.5) and (4.7) into (4.4) leads to

$$\Delta y_{it} = (\rho_t - 1)y_{it-1} + Z_i\tilde{\gamma}_t + \sum_l \sum_m st(l,m)\pi_{lm} + ((\alpha+1)\eta_{it} + (\theta-\rho)\zeta_{it-1} + \omega_{it}) \tag{4.10}$$

[14] Similar results with related measurement error models have been derived in Fields *et al.* (2003a) and Antman and Mckenzie (2005a).

[15] It can be proved that this estimation will not be biased even under a more general measurement error structure in which the measurement error is correlated with both the transitory *and* the permanent components of earnings.

It is clear from this expression that the variable y_{it-1} is not independent from the error term, because of its correlation with ζ_{it-1}. This correlation will bias *all* the parameter estimates in the model, when estimated by OLS. The only exception to this would occur if $\theta = \rho$, i.e., if the correlation in transitory earnings equals the correlation in the idiosyncratic measurement error, something which seems extremely unlikely in practice.

The route taken when estimating eqn. (4.10) will be to use occupational dummies together with wealth proxies as instruments for y_{it-1} (in addition to all the variables included under Z and X). This method is not fully satisfactory, since in such IV estimation there will be a component of transitory mobility that will not be captured by the instruments, leading to underestimation of the conditional mobility in the model.

4.3.5 Attrition Bias

Concerning attrition and non-reporting of the earnings variable in the panel, the approach taken here is to abandon the pretension of obtaining precise point estimates of $E(\Delta y_{it}|y_{it-1})$, and instead turn to partial identification techniques (see for instance Manski, 1995, 2003). The idea underlying partial identification analysis is to provide a whole region where the conditional expectation of interest can lie, given that the information available is not complete. The advantage of this approach is that the set of assumptions made in generating the identification regions are minimal. In this chapter partial identification analysis will be applied to the case of missing outcomes (earnings in the final period) due to attrition and non-reporting. These are the two biggest sources for missing data in the sample.

In order to see how these methods work in the present context, let z_i be an indicator variable on whether the earnings of individual i are observed in the final period. Also for simplicity of exposition, rescale $E(\Delta y_{it}|y_{it-1})$ to make it lie inside the [0,1] interval. By the law of iterated expectations

$$\begin{aligned} E(\Delta y_{it}|y_{it-1}) = E(\Delta y_{it}|y_{it-1}, z_i = 1) \cdot P(z_i = 1|y_{it-1}) \\ + E(\Delta y_{it}|y_{it-1}, z_i = 0) \cdot P(z_i = 0|y_{it-1}). \end{aligned}$$

The data alone reveals $E(\Delta y_{it}|y_{it-1}, z_i = 1)$, $P(z_i = 1|y_{it-1})$ and $P(z_i = 0|y_{it-1})$ only. Since the whole expectation was rescaled to lie between [0,1], it follows that the lowest value $E(\Delta y_{it}|y_{it-1})$ can take is

$$E(\Delta y_{it}|y_{it-1}, z_i = 1) \cdot P(z_i = 1|y_{it-1})$$

when $E(\Delta y_{it}|y_{it-1}, z_i = 0) = 0$, and the highest value it can take is

$$E(\Delta y_{it}|y_{it-1}, z_i = 1) \cdot P(z_i = 1|y_{it-1}) + P(z_i = 0|y_{it-1})$$

when $E(\Delta y_{it}|y_{it-1}, z_i = 0) = 1$. Any point between these two bounds forms the identification region $H[E(\Delta y_{it}|y_{it-1})]$. Any values inside this region are logically possible for $E(\Delta y_{it}|y_{it-1})$, given the amount of attrition and non-response in the data. This identification region is estimated by nonparametric methods.

It is important to remark that without any extra assumptions, the information contained in $H[E(\Delta y_{it}|y_{it-1})]$ is all that the data reveals about $E(\Delta y_{it}|y_{it-1})$. Hence, the estimation of this region without further assumptions gives the worst case scenario for the impact of the loss of information due to attrition and non-reporting. Plausible extra assumptions can narrow the width of the identification region $H[E(\Delta y_{it}|y_{it-1})]$. However, that refinement will be left for further extensions of this exercise.

4.4 Results

4.4.1 Unconditional Mobility

This section presents the results that pertain to the unconditional mobility analysis as described in section 4.3.1. In particular, it presents the estimates for the parameter β_1 from equation (4.1), i.e.,

$$\Delta y_{it} = \beta_0 + \beta_1 y_{it-1} + u_{it}.$$

Figure 4.1 plots β_1 obtained by Least Squares, both for earnings in levels and logarithms. The graphs show unanimously that there is convergence between the earnings of rich and poor. That is, over a calendar year the initially poor got richer, and the initially rich got poorer.[16] The graphs of the parameters do not show a specific trend or pattern of this convergence over time.[17]

Since the LS estimates can be strongly affected by the presence of outliers, a more robust specification is presented in Figure 4.2 which plots the estimates of β_1, now computed by a median regression. These estimates confirm the convergence result obtained through LS. In the case of the regression with earnings in levels the estimates present an inverted U-pattern in the second half of the sample. The estimates for log-earnings are quite similar to the LS ones, presumably because the logarithmic transformation diminishes the effects of outliers in the LS regression.

Although the finding of convergence in earnings between rich and poor over a calendar year seems clear (leaving aside issues of measurement error), it is not

[16]It is important to remark that this convergence in earnings is not just the result of the inclusion of unemployed individuals in the sample. The fraction of unemployed individuals in the sample is too small to drive this result, and estimations of β_1 excluding the unemployed give undistinguishable parameter estimates, as Figure 4.18 in the Appendix demonstrates.

[17]If anything, the graph of log-earnings has an increasing concave shape.

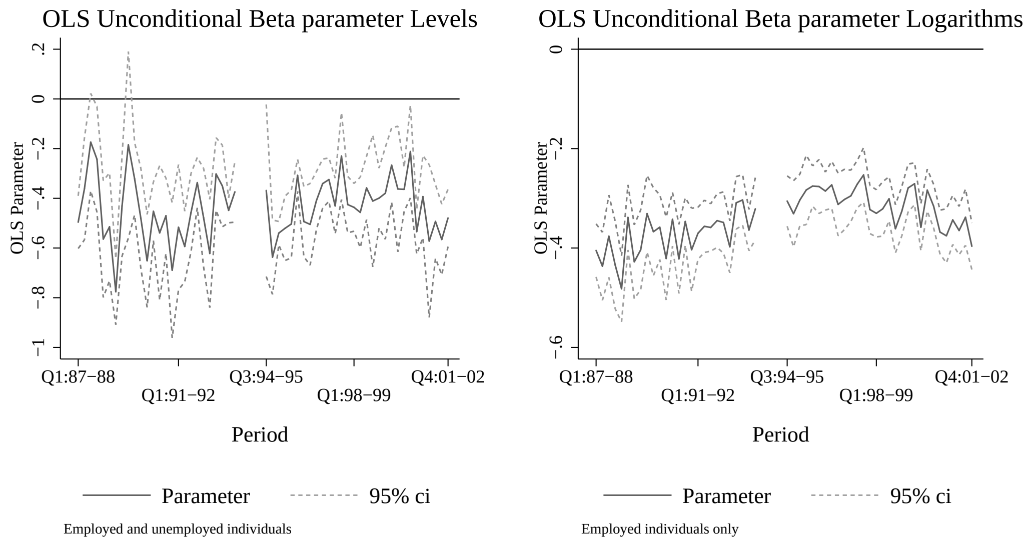


Figure 4.1: OLS Unconditional Mobility Parameter

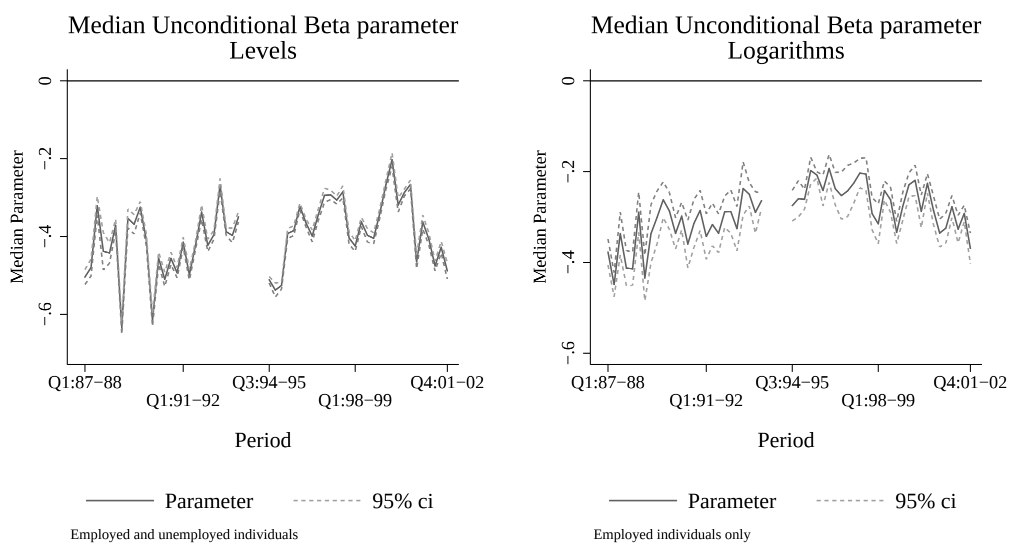


Figure 4.2: Unconditional Mobility Parameter. Median Regression

evident how to interpret this result. On one side it could mean that low-earners are catching up with high-earners; but, as previously mentioned, this could also be the product of an adjustment to a temporary shock in earnings, without any true approaching between rich and poor. Evidence supporting this second interpretation is presented in Figure 4.3.

The graphs in Figure 4.3 plot the average earnings profiles for individuals classified at different points in time into quintiles of the earnings distribution. They show that in the quarter in which the quintile classification takes place, the earnings of the individuals in the lowest quintile are considerably lower than at any other period. Similarly, in this period the average earnings of the individuals in the top quintile appear to be considerably larger than what they usually are. In other words, classifying individuals as rich and poor based on the earnings of a single period exacerbates their apparent advantage or deprivation (depending on the case). The implication of this finding for the present unconditional mobility estimations is that when regressing Δy_{it} on initial earnings, part of the convergence obtained reflects the adjustment of earnings back to their "regular" level, and not necessarily convergence between these earnings profiles.[18]

In order to capture the relationship between mobility and "permanent" advantage, regressions like (4.1) are estimated now using measures of permanent advantage as a regressor. In particular, two measures of permanent advantage are considered. The first one is average earnings over time, for each individual. The second one is constructed by regressing initial earnings on a set of variables that are related to permanent advantage like age, education, gender, and wealth proxies, and uses the parameter estimates of this regression to predict $\hat{y}_{it-1}$. The

[18] Although this graph corresponds to one of the last panels in the sample, similar plots for other years lead to the same conclusion.

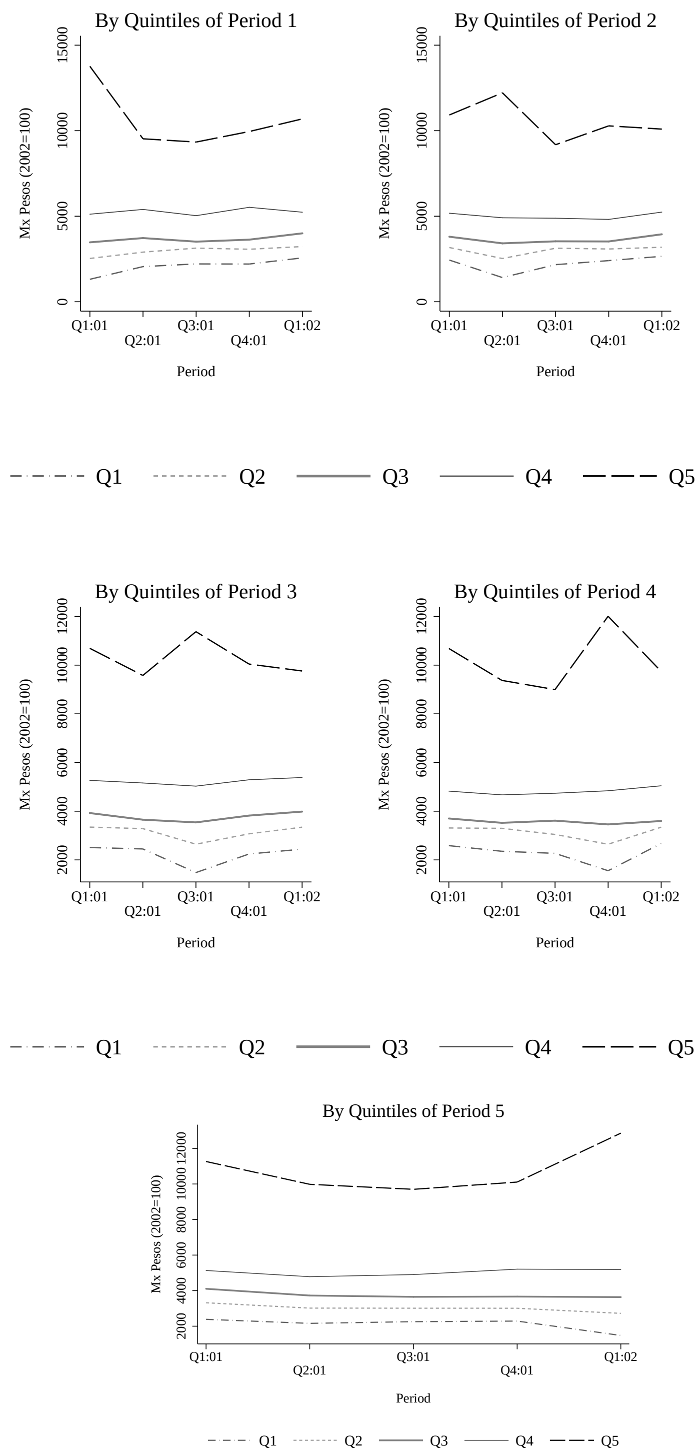


Figure 4.3: Earnings Profiles by Quintile Groups Classified at Different Periods

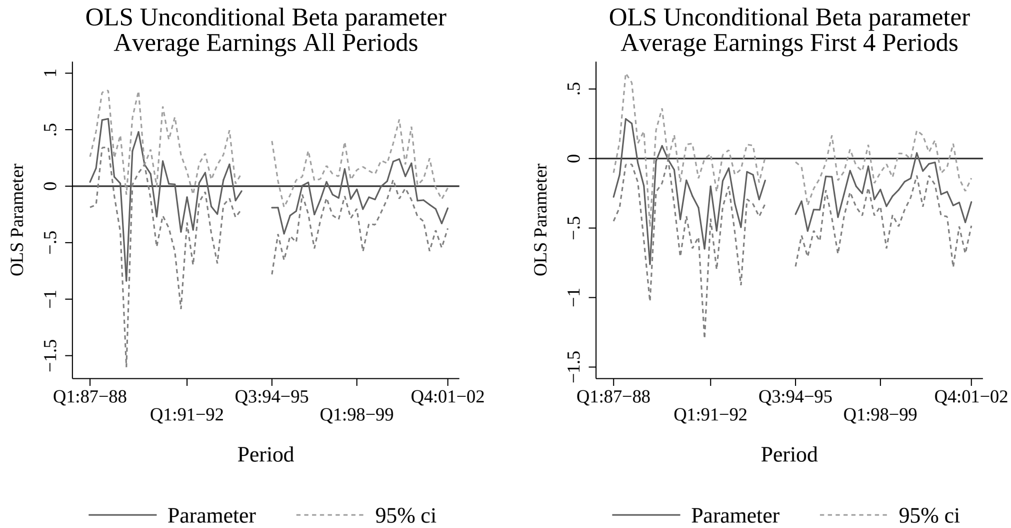


Figure 4.4: Unconditional Mobility Parameter with Average Earnings as Measure of Permanent Advantage

results of these estimations are plotted in Figures 4.4 to 4.6.[19]

Figure 4.4 contains the graphs where average earnings is a measure of permanent advantage. Here two exercises are performed, one averaging earnings in the first 4 quarters and the second averaging earnings over the full 5 quarters of observations. Both estimations are performed only for earnings in levels. Figure 4.5 contains the estimates when the measure of permanent advantage is the predicted earnings stemming from a regression of initial earnings on human capital variables and wealth proxies. These predictions are made both for earnings in levels and log-earnings. Finally, Figure 4.6 plots the results corresponding to a predicted $\hat{y}_{it-1}$ coming from a regression that, in addition to all the regressors

[19] In the LS case this just amounts to an instrumental variable regression.

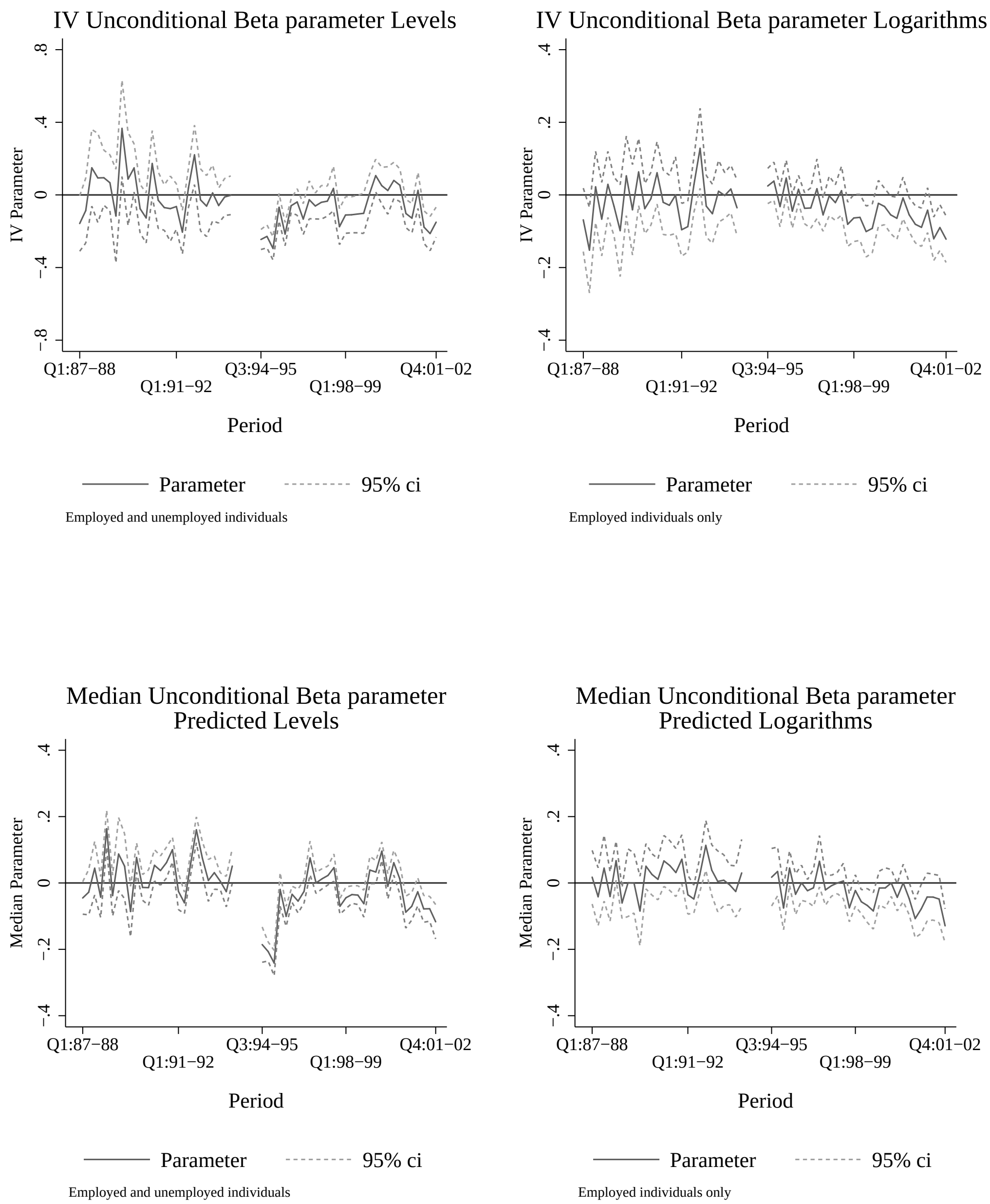


Figure 4.5: Unconditional Mobility Parameter with Predicted Earnings as a Measure of Permanent Advantage. Human Capital and Wealth Proxies Controls.

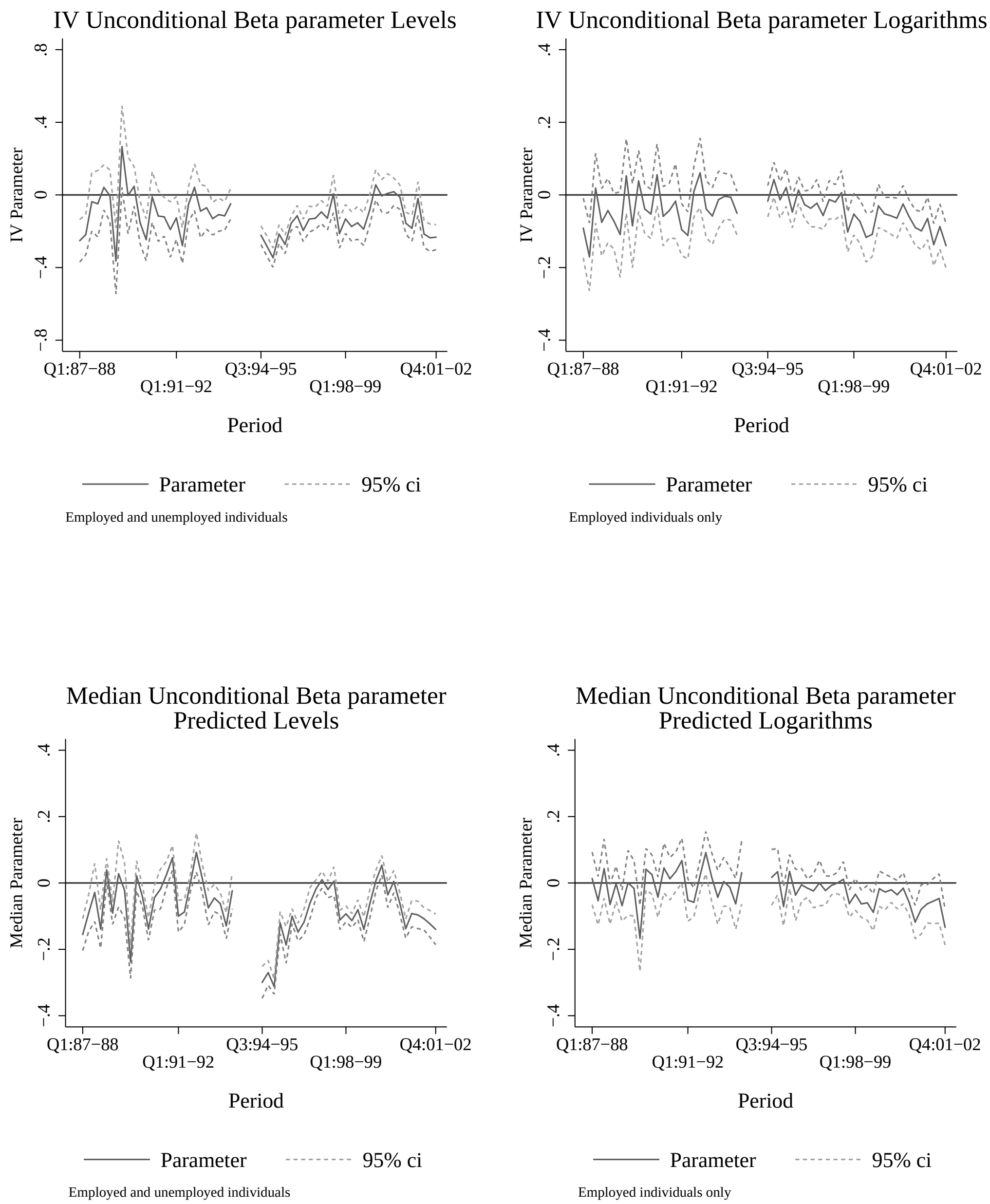


Figure 4.6: Unconditional Mobility Parameter with Predicted Earnings as Measure of Permanent Advantage. Human Capital, Wealth Proxies and Sector Controls.

previously mentioned, also includes sectoral dummies indicating whether the individual is a formal wage worker, informal wage worker, formal self-employed or informal self-employed.

All these figures show that yearly mobility is unrelated to the generated measures of permanent advantage for most of the years in the sample. The only major exception occurs right after the 1994 Peso crisis in the regressions with earnings in levels when there is convergence in earnings. Besides this episode, only a couple of divergent results in the early nineties, and a slightly convergent pattern during 1998 and 2001 are observed. One interesting point is that the lack of *log*-convergence after the Peso crisis implies that in this episode the "permanent" rich lost more than the "permanent" poor, but their losses were proportional to their higher levels of earnings.

In order to have a better idea of the first-stage predictions in the previous exercises, Table 4.1 reproduces one of these first-stage regressions (for the 1st-quarter of 1987) for the variable of earnings in levels. This regression shows that earnings are positively related to age, education (with an increasing convex pattern), being male and to cluster average earnings (a wealth proxy). Being a formal self-employed has the largest returns followed by informal self-employment, formal wage work and informal wage work.

The R^2-Adjusted for all the regressions over time, both in levels and logarithms, are presented in Figure 4.7. This figure shows that the R^2 of these first-stage regressions fluctuates between 25% and 45%, with the logarithmic regressions having a better fit. These magnitudes are in line with similar earnings equations estimated for other countries. If the assumed model of earnings is right, then at a given point in time, approximately 30% of the variation in earnings is due to differences

Table 4.1: First-Stage of IV Prediction. Q1:87. Dep.Var.:Earnings

	Coef./(s.e.)	
Age		
Linear	140.385	**
	(62.76)	
Squared	-1.298	
	(0.84)	
Years of Education		
Linear	63.693	
	(51.73)	
Squared	8.597	***
	(2.69)	
Male	1248.911	***
	(167.04)	
Cluster Average Earnings	0.745	***
	(0.11)	
Sector of Employment		
Informal Worker	3699.613	***
	(441.63)	
Informal Self-employed	4416.127	***
	(331.24)	
Formal Self-employed	7546.767	***
	(1399.52)	
Formal Worker	4289.504	***
	(349.13)	
Constant	-8224.471	***
	(1209.16)	
R2-Adj.	0.297	
N	4063	

* p < 0.1, ** p < 0.05, *** p < 0.01

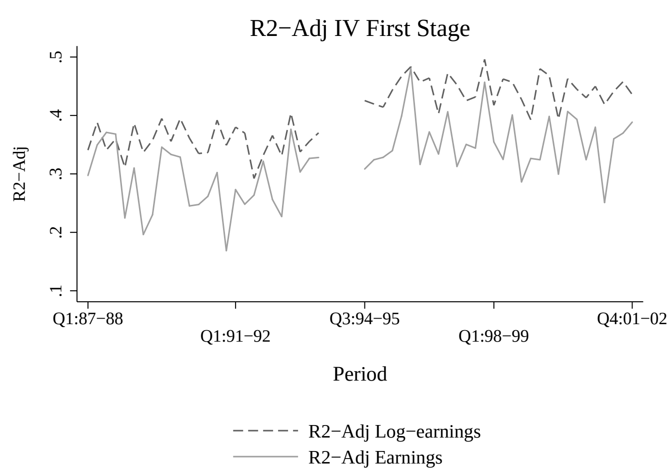


Figure 4.7: R^2 Adjusted of First Stage of IV Model. Human Capital, Wealth Proxies, and Sector Controls.

in permanent earnings. Again, the caveat raised in footnote 7 about the effects of unobserved characteristics applies here.

Summarizing, these findings confirm that the strong convergence obtained when using reported earnings as a measure of initial advantage was mostly due to a short-run adjustment of earnings back to their permanent level. In other words, the mobility over a year did not alter the permanent advantage of the individuals in the economy. The only exception to this occurred in the aftermath of the 1994 Peso crisis. This crisis brought proportional losses to everybody in the economy, making the richer individuals lose more than anybody else in absolute terms.

4.4.2 Conditional Mobility and the Determinants of Earnings Changes

To start the presentation of the results on conditional mobility the parameter $\rho_t - 1$ from eq. (4.4), i.e.,

$$\Delta y_{it} = (\rho_t - 1) y_{it-1} + Z_i \tilde{\gamma}_t + \sum_l \sum_m st(l,m)\pi_{lm} + \eta_{it}$$

is plotted for several specifications.

Figure 4.8 shows this parameter when the included controls are only time-invariant variables Z_i that capture human capital characteristics like age, education and gender, plus regional control dummies. As it can be appreciated, there is always convergence to the conditional mean, and this convergence is slightly stronger than the unconditional convergence presented in the previous section. This means that the overall effect of the human capital and regional controls is to generate divergence in earnings, so that once these socioeconomic variables are explicitly accounted for, the convergence in earnings is stronger. Also it can be

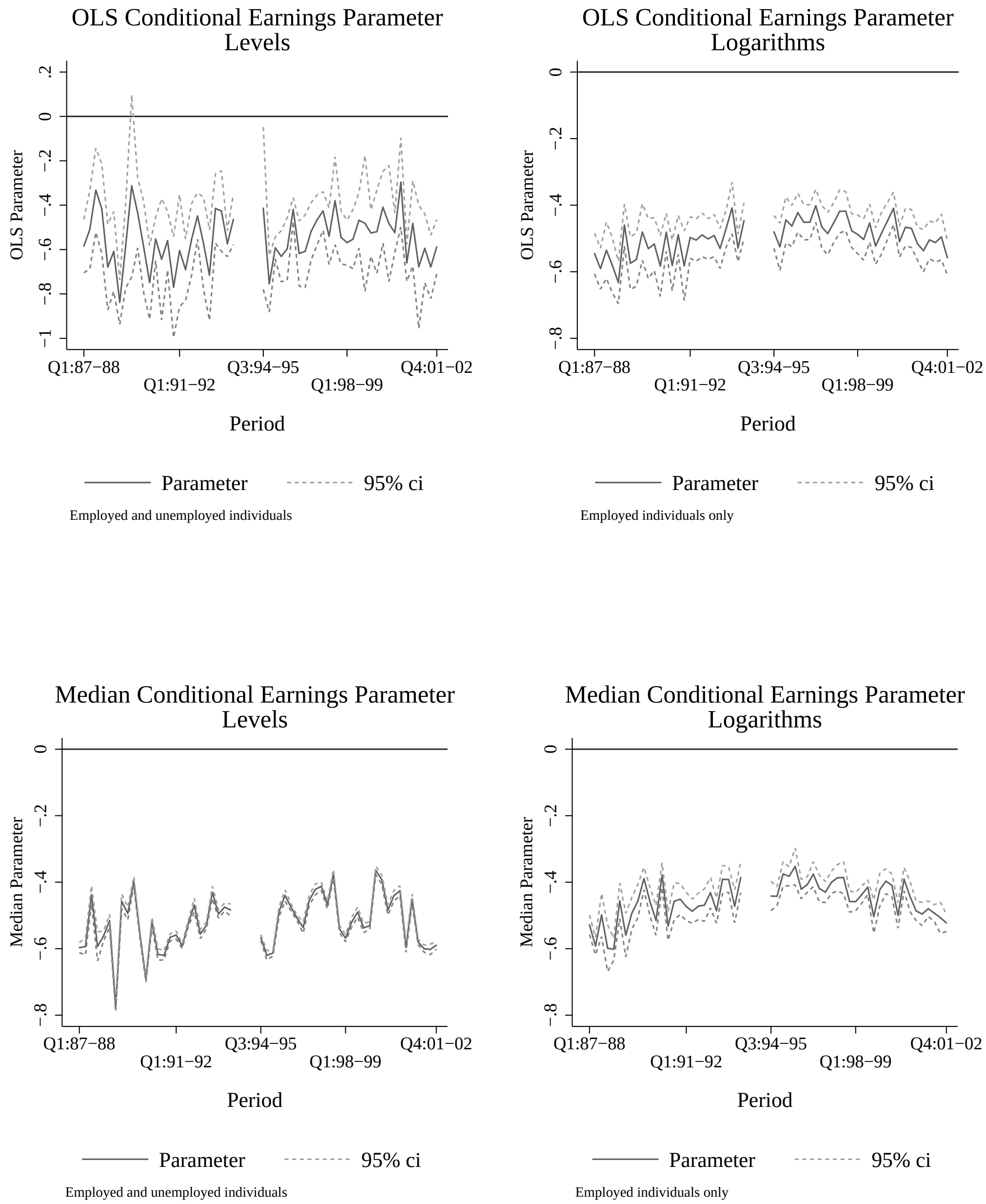


Figure 4.8: Conditional Mobility Parameter. Human Capital and Regional Controls.

seen that this parameter, which is around -0.6 would imply a value for ρ of about 0.4, i.e., the auto-regression parameter in transitory earnings is about 0.4.

These estimates are complemented by estimating the same conditional convergence parameter now including sector transition dummies as additional controls. As it can be seen from Figure 4.9 the conditional mobility parameters are very similar to the ones of Figure 4.8, hence, sector transition controls do not seem to affect much the conditional convergence rates.

Before moving to analyze the specific effect of the Z_i, X_{it} variables on earnings mobility, the results from estimating different $\rho_t - 1$ by subgroups of the population are presented. The reason for doing this extension is that the rate at which earnings converge to the conditional mean might differ for different groups of the population. These estimations include interactions by age, education, gender, sector, and region groups.[20] The particular form of the equations estimated is

$$\Delta y_{it} = \sum_g \left[\left((\rho_t^g - 1) y_{it-1} + Z_i \tilde{\gamma}_t^g + \sum_l \sum_m st(l,m) \pi_{lm}^g \right) \mathbb{1}(i \in g) \right] + \eta_{it}$$

where g denotes group, and $\mathbb{1}(i \in g)$ is an indicator variable that takes value 1 if the individual i belongs to group g, and zero otherwise. Notice that because $\tilde{\gamma}_t$ and π_{lm} are both functions of ρ_t, the model that allows ρ_t to vary by groups implies a fully interactive conditional mobility equation.

Figures 4.10-4.14 contain the estimates of these interactions. Out of all the interactions estimated, the groups that seem to have some noticeable differences in their convergence rates are the education and sector groups. In the case of education groups, the higher the education the smaller the convergence rates. This means that shocks are more persistent for more educated individuals. In the case of

[20]For a definition on how the age and education groups were defined see Chapter 3.

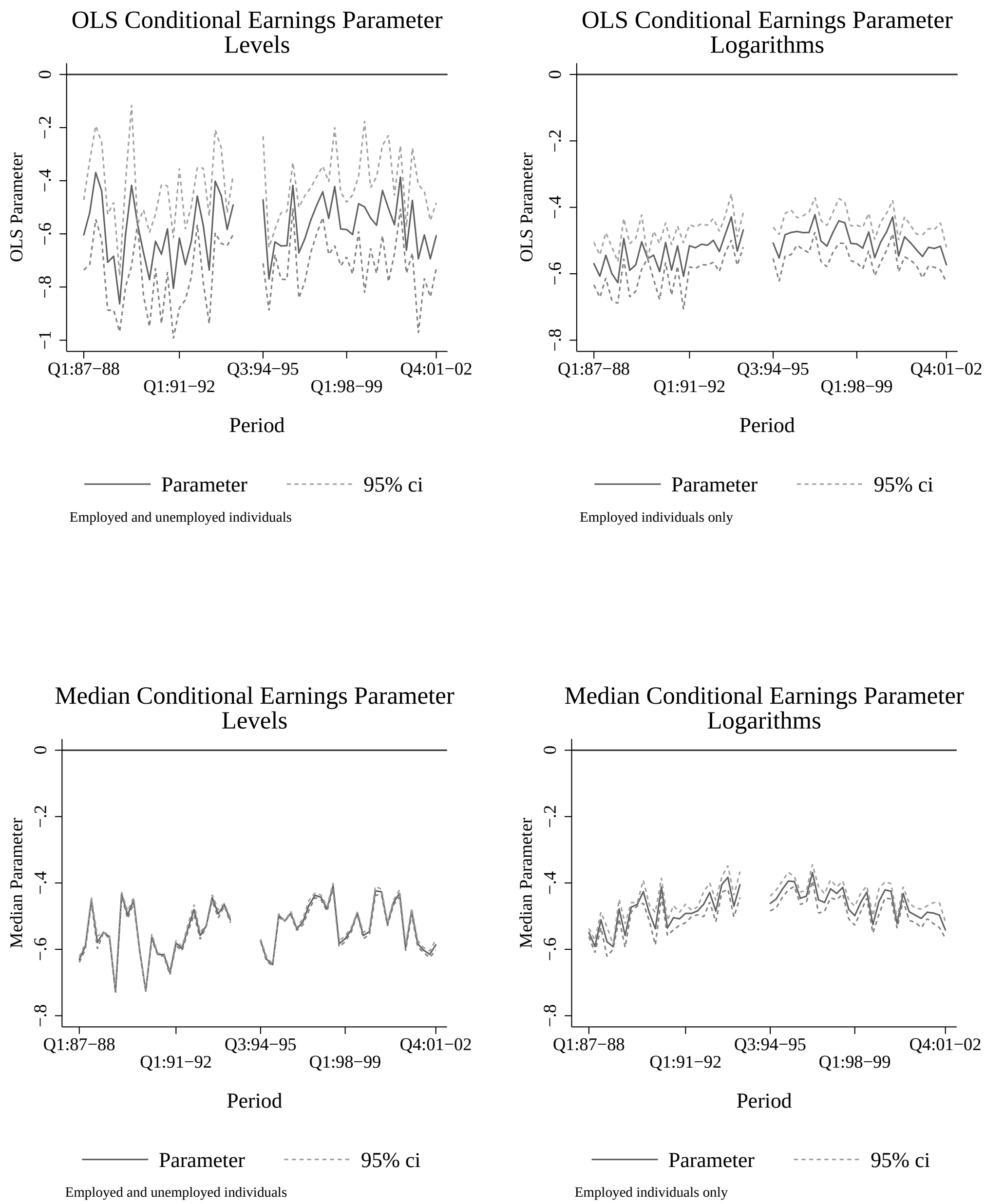


Figure 4.9: Conditional Mobility Parameter. Human Capital, Regional and Sector Transition Controls.

sector groups, the formal wage workers present the lowest convergence rates, while the formal self-employed exhibit the highest ones, although their convergence rates also fluctuate more.

The only other noticeable differences in these graphs is that during the second half of the sample women tend to have lower convergence rates than men, but this result is only valid in the logarithmic specification, and is statistically significant only for a couple of years. In the case of regional interactions, the differences in $\rho_t - 1$ are statistically significant in about half of the cases, but it is not clear what pattern emerges.

In order to present the direct effects of socioeconomic variables on earnings changes, the data is first pooled for all the years in the sample. The results from these regressions are shown in Table 4.2.

The results for the regression in levels show that age increases mobility, but at a decreasing rate. Education on the contrary has a convex pattern. From 0 to about 3 years of education an extra year of education reduces earnings mobility, but after that point it increases it.[21] Being a male has a large positive impact on mobility.

Out of all the sector transitions, the one into formal self-employment is always associated with the largest gains, after controlling for everything else. This could be generated by the potential inclusion of capital gains in the earnings reported by the self-employed. Aside from movements into unemployment (which trivially involve losses), the most negative conditional mobility is associated with transitions into informal wage work. The other destination sectors (Formal wage work and informal self-employment) are between the two previous extremes, and they have

[21]This effects is statistically significant only in the specification in levels.

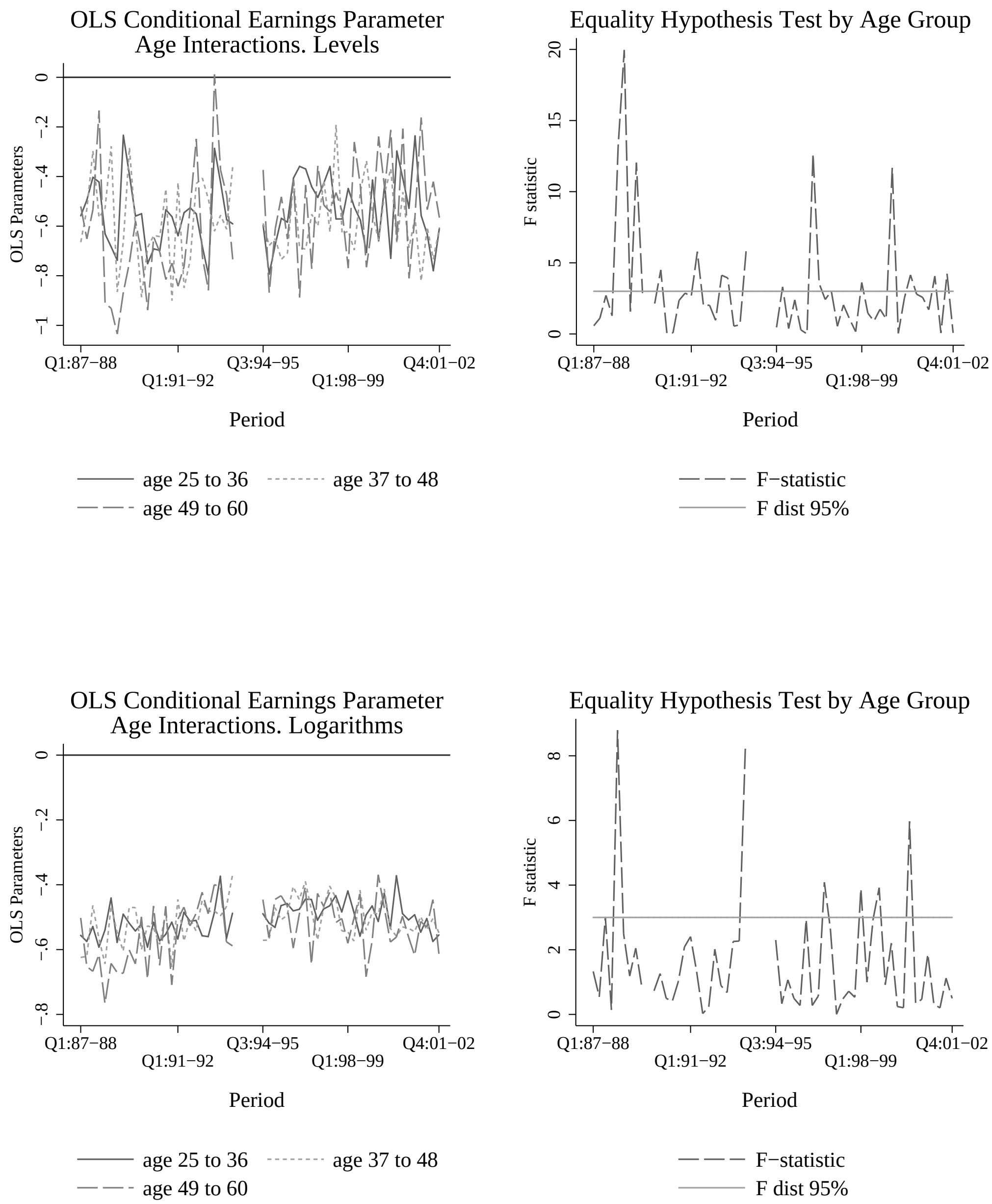


Figure 4.10: Conditional Mobility Parameter by Age Group.

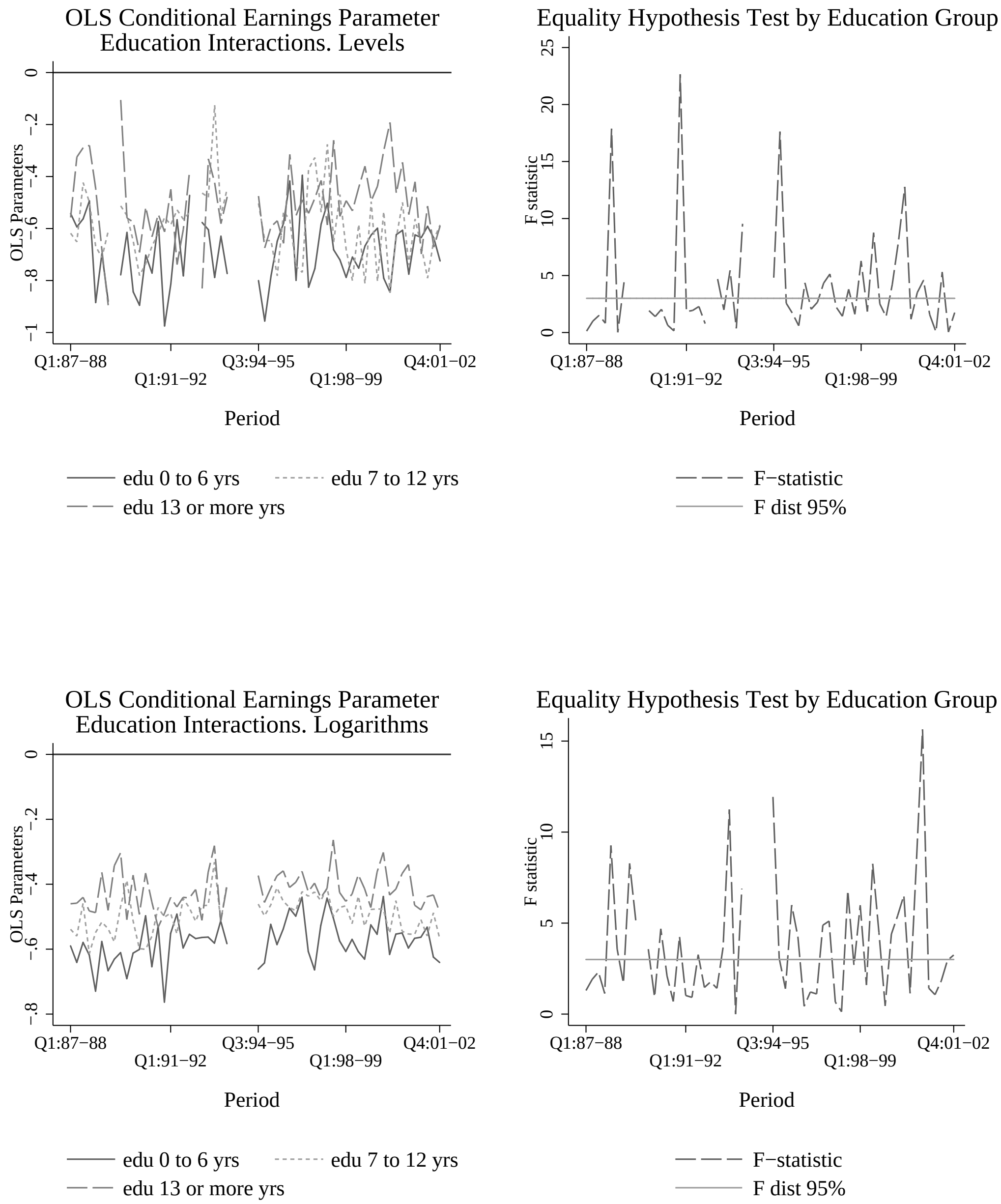


Figure 4.11: Conditional Mobility Parameter by Education Group.

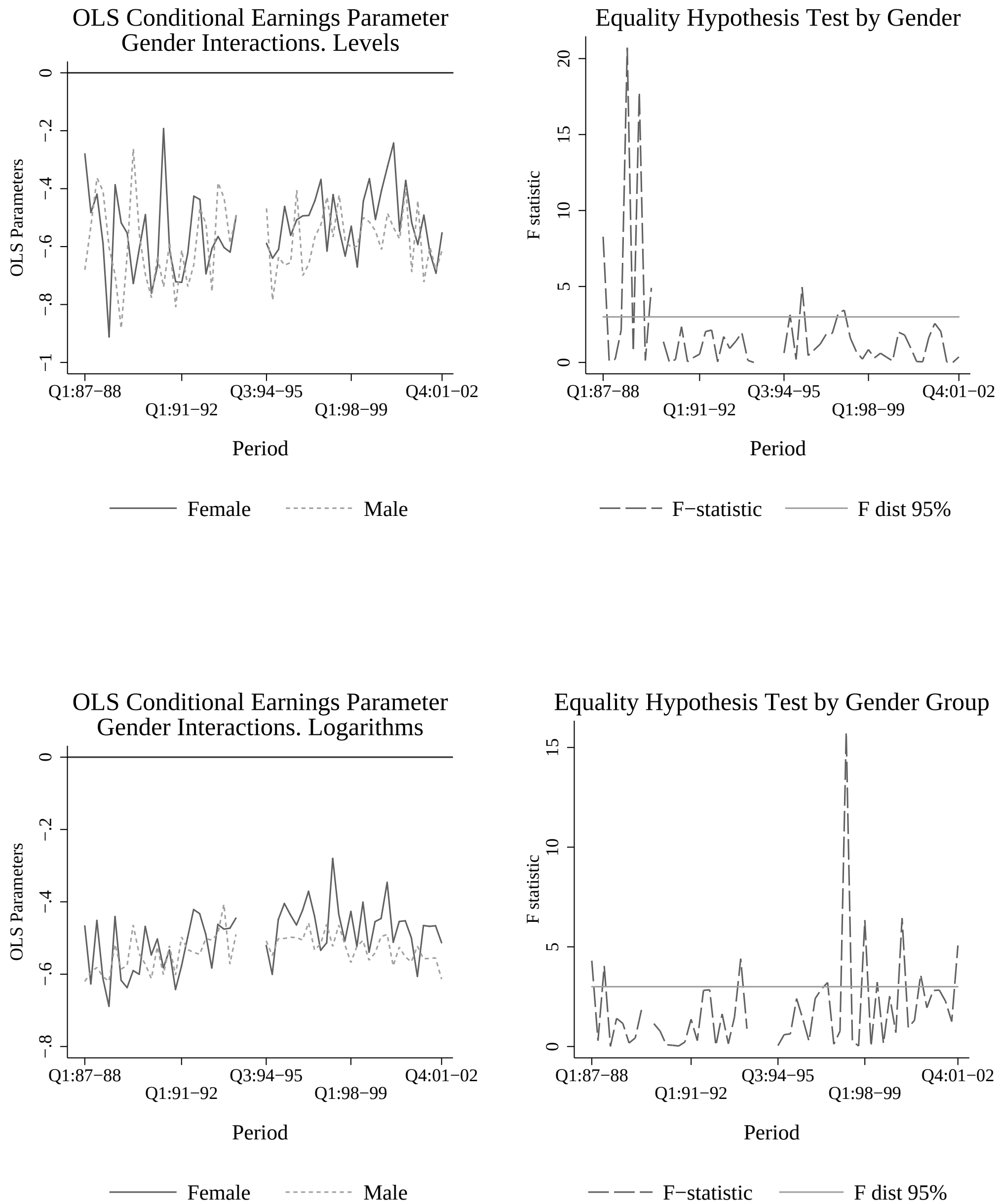


Figure 4.12: Conditional Mobility Parameter by Gender.

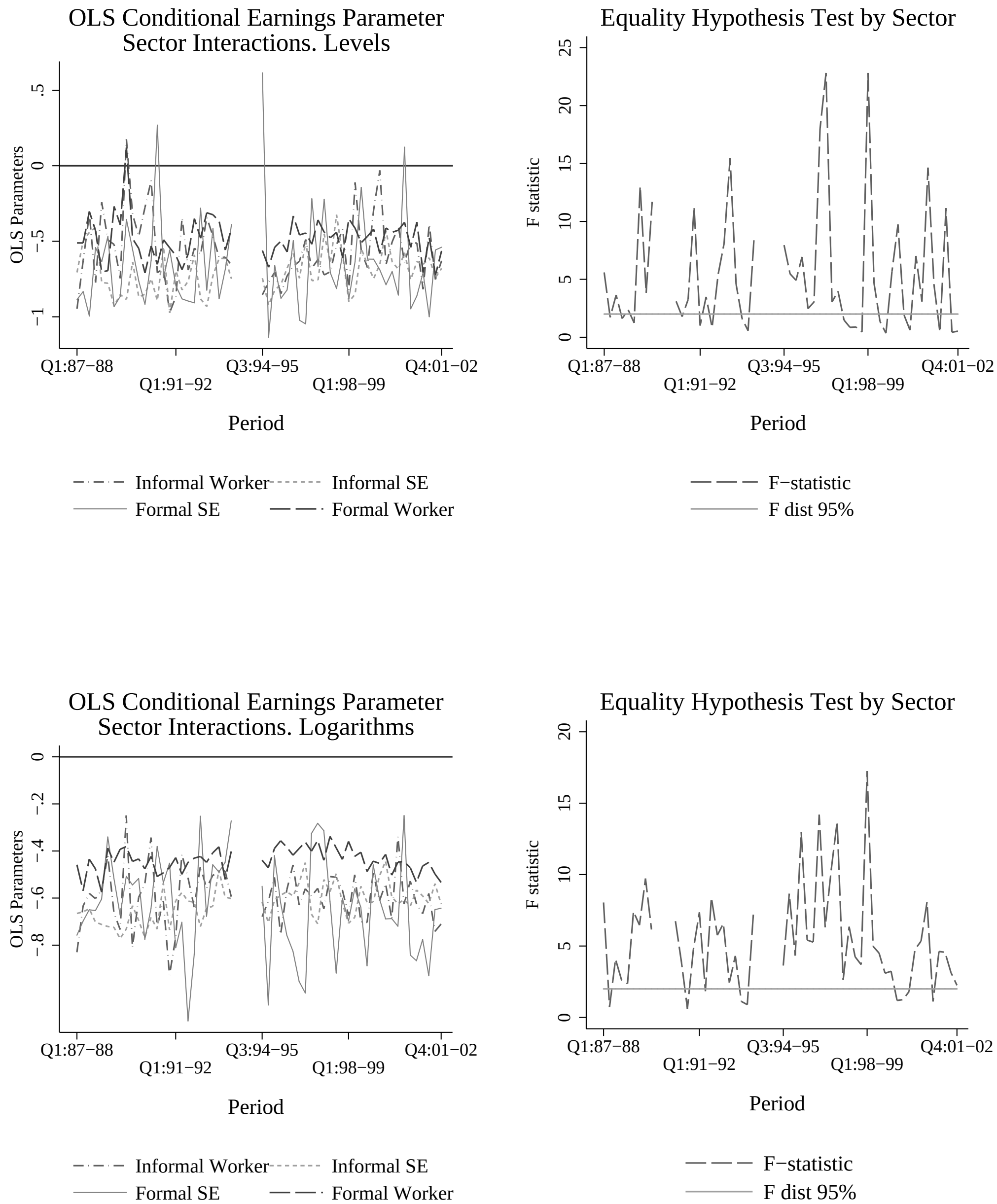


Figure 4.13: Conditional Mobility Parameter by Sector.

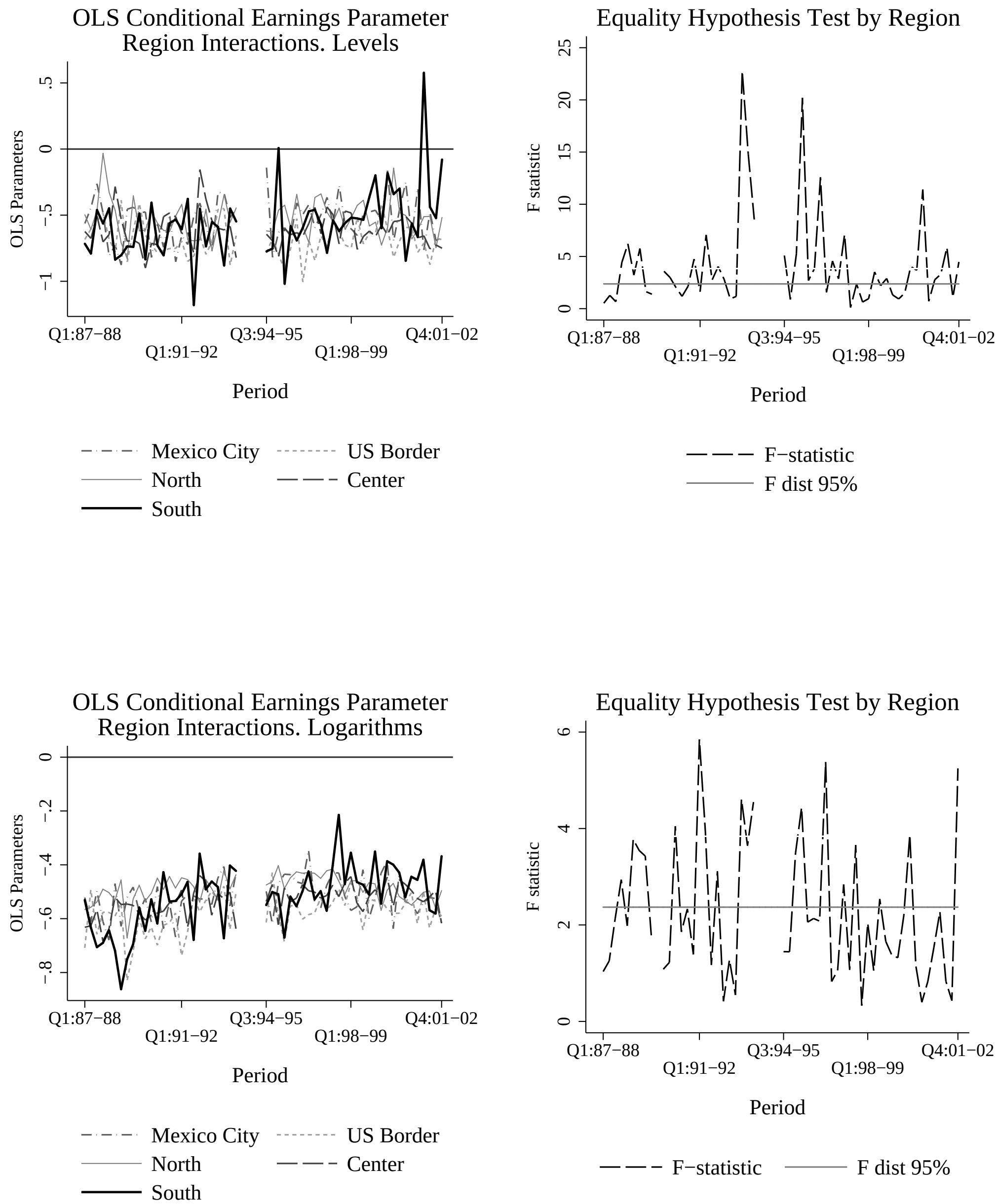


Figure 4.14: Conditional Mobility Parameter by Region.

roughly equal mobility.[22]

Finally, in what concerns the regional analysis the cities along the US Border and in the North experience higher conditional mobility, while the Center and the South exhibit less positive conditional mobility. The omitted region is Mexico City. The results for the regression in logarithms show a similar pattern to the one above described.

The set of estimations of eq. (4.4) broken down by pooled subperiods are shown in Tables 4.3 and 4.4.

In these tables, the effect of age is similar to the one obtained in the pooled regressions; however, the positive effect of age has diminished over time. Again, the convex pattern is found for the years of education variable. However, as time goes by the inflection point at which the positive effect on mobility kicks in is located at higher levels of education, i.e., the negative effect of education on mobility for individuals with low education has become more pervasive.[23]

The positive effects of being male on mobility are large in all periods, and under all the specifications. However, while in the specification in levels the effect appears to follow the cycle of average earnings (highest in the 87-93 period, lowest in the 95-99, otherwise in the middle); in the specification in logarithms this positive effect has become stronger over the years.

[22]It is interesting to recall that a similar ranking of mobility by sector of destination was obtained in the previous chapter, where Directional mobility by sector transition was studied.

[23]A similar conclusion applies to the regression in logarithms, while in the first period going from 1987 to 1993 education had an increasing convex effect on log-mobility, after 1999 a negative effect for low educated people appeared.

Table 4.2: Pooled OLS Regression.
Dep. Var.: Change in Reported Earnings

	Earnings Change		Log-Earnings Change	
	Coef./(s.e.)		Coef./(s.e.)	
Initial Earnings	-0.593	***		
	(0.02)			
Initial Log-Earnings			-0.501	***
			(0.00)	
Age				
Linear	120.838	***	0.020	***
	(16.95)		(0.00)	
Squared	-1.263	***	-0.000	***
	(0.21)		(0.00)	
Education				
Linear	-142.341	***	-0.001	
	(13.35)		(0.00)	
Squared	19.311	***	0.002	***
	(0.90)		(0.00)	
Male	802.052	***	0.148	***
	(32.75)		(0.00)	
Sector Transitions				
Unemployed to Informal Worker	3030.169	***		
	(131.97)			
Unemployed to Informal Self-employed	3979.493	***		
	(224.04)			
Unemployed to Formal Self-employed	19346.141	***		
	(3964.50)			
Unemployed to Formal Worker	4062.554	***		
	(145.94)			
Informal Worker to Unemployed	-166.703			
	(146.01)			
Informal Worker to Informal Worker	2480.740	***		
	(116.57)			
Informal Worker to Informal Self-employed	3212.336	***	0.188	***
	(130.29)		(0.02)	
Informal Worker to Formal Self-employed	7293.322	***	0.900	***
	(912.41)		(0.14)	
Informal Worker to Formal Worker	2850.443	***	0.166	***
	(122.14)		(0.01)	
Informal Self-employed to Unemployed	-1204.665	***		
	(192.02)			
Informal Self-employed to Informal Worker	2006.394	***	-0.104	***
	(124.17)		(0.02)	
Informal Self-employed to Informal Self-employed	3124.402	***	0.119	***
	(133.73)		(0.01)	
Informal Self-employed to Formal Self-employed	8498.560	***	0.640	***
	(766.89)		(0.05)	
Informal Self-employed to Formal Worker	2602.279	***	0.139	***
	(145.01)		(0.02)	
Formal Self-employed to Unemployed	-5689.423	***		
	(767.25)			
Formal Self-employed to Informal Worker	1591.859	*	-0.144	
	(906.45)		(0.14)	
Formal Self-employed to Informal Self-employed	4582.217	***	0.266	***
	(561.97)		(0.04)	
Formal Self-employed to Formal Self-employed	11571.363	***	0.690	***
	(917.95)		(0.04)	
Formal Self-employed to Formal Worker	4884.640	***	0.271	***
	(740.12)		(0.05)	
Formal Worker to Unemployed	-1911.110	***		
	(176.64)			
Formal Worker to Informal Worker	2456.328	***	0.003	
	(121.61)		(0.01)	

Table 4.2 (Continued)

Formal Worker to Informal Self-employed	3461.882	***	0.160	***
	(160.32)		(0.02)	
Formal Worker to Formal Self-employed	8579.295	***	0.630	***
	(824.12)		(0.05)	
Formal Worker to Formal Worker	3116.878	***	0.192	***
	(123.98)		(0.01)	
Region				
Mexico City (omitted)				
US Border	562.339	***	0.129	***
	(53.01)		(0.01)	
North	112.688	**	0.024	***
	(45.09)		(0.01)	
Center	-111.184	***	0.003	
	(35.20)		(0.01)	
South	-461.877	***	-0.085	***
	(51.45)		(0.01)	
Constant	-4402.654	***	3.204	***
	(354.63)		(0.05)	
R-squared	0.339		0.262	
N	236854		229111	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The effects of sector transitions on mobility are similar to the ones described for the pooled case. Transitions into informal wage work are associated with the most negative mobility, while transitions into formal self-employment are associated with the largest positive gains. In general, the transitions into informal self-employment bring more upward conditional mobility than the ones into formal wage work, but this effect is sometimes reversed during the aftermath of the 1994 Peso crisis.

It is important to stress here that these parameter estimates of sector transitions should not to be taken as evidence of segmentation between formal and informal sectors in Mexico. These parameters just reflect the conditional earnings changes experienced by movers and stayers, and they need not reflect the counterfactual gains a randomly selected individual would experience by moving from sector x into sector y.

Regarding the regional patterns the effects are similar to the ones described in the pooled case. It is worth mentioning that the positive effect on mobility of living in a US Border city has become stronger over time. Whether this is related to the increasing activity of "maquiladoras" (American assembly factories that benefit from the comparatively cheap labor across the border) is something that requires further study.

So far the results presented have assumed that the earnings variables are measured without error, or more precisely that they are correctly reported. It also assumed that the individuals that disappear from the sample or that do not report their earnings are doing so at random. Both assumptions are unrealistic and require further scrutiny. These issues are tackled in the next sections.

Table 4.3: OLS Regression by Periods. Levels. Dep. Var.: Change in Reported Earnings

	Q1:87-Q2:94 Coef./(s.e.)		Q3:94-Q1:99 Coef./(s.e.)		Q2:99-Q4:02 Coef./(s.e.)	
Initial Earnings	-0.621	***	-0.571	***	-0.587	***
	(0.03)		(0.03)		(0.04)	
Age						
Linear	194.290	***	92.974	***	66.527	**
	(29.20)		(25.14)		(32.39)	
Squared	-2.099	***	-0.928	***	-0.613	
	(0.36)		(0.32)		(0.41)	
Education						
Linear	-91.764	***	-148.619	***	-181.971	***
	(21.02)		(22.76)		(28.00)	
Squared	19.135	***	18.146	***	21.001	***
	(1.38)		(1.52)		(1.87)	
Male	968.224	***	597.598	***	780.774	***
	(54.73)		(43.76)		(71.22)	
Sector Transitions						
Unemployed to Informal Worker	4111.114	***	2499.281	***	3030.385	***
	(272.40)		(143.01)		(280.08)	
Unemployed to Informal Self-employed	5060.229	***	3045.929	***	5299.060	***
	(393.15)		(240.27)		(767.00)	
Unemployed to Formal Self-employed	14615.738	*	22656.096	***	7423.921	***
	(7786.13)		(3891.63)		(394.66)	
Unemployed to Formal Worker	4237.511	***	3695.184	***	4785.051	***
	(268.05)		(207.29)		(364.79)	
Informal Worker to Unemployed	-8.605		-379.147	*	265.666	
	(288.87)		(203.14)		(282.29)	
Informal Worker to Informal Worker	3104.463	***	1957.628	***	2824.087	***
	(257.70)		(143.53)		(265.59)	
Informal Worker to Informal Self-employed	4017.568	***	2533.454	***	3384.354	***
	(281.09)		(164.19)		(277.84)	
Informal Worker to Formal Self-employed	7408.207	***	7122.204	***	8136.863	***
	(1112.99)		(2171.75)		(2585.46)	
Informal Worker to Formal Worker	3484.269	***	2199.401	***	3143.540	***
	(262.05)		(146.14)		(275.63)	
Informal Self-employed to Unemployed	-1213.697	***	-1206.367	***	-939.647	**
	(369.46)		(260.78)		(424.68)	
Informal Self-employed to Informal Worker	2607.156	***	1474.242	***	2260.800	***
	(257.33)		(165.95)		(292.77)	
Informal Self-employed to Informal Self-employed	3905.512	***	2368.119	***	3319.076	***
	(272.77)		(175.49)		(297.95)	
Informal Self-employed to Formal Self-employed	9982.576	***	6168.406	***	8933.459	***
	(1436.25)		(1189.88)		(1020.36)	
Informal Self-employed to Formal Worker	2969.937	***	1977.717	***	3138.862	***
	(289.38)		(195.10)		(316.64)	
Formal Self-employed to Unemployed	-7776.634	***	-4720.922	***	-3698.514	***
	(855.26)		(736.62)		(507.12)	
Formal Self-employed to Informal Worker	2590.661	*	448.713		807.239	
	(1404.30)		(675.60)		(1852.53)	
Formal Self-employed to Informal Self-employed	5802.505	***	2658.578	***	4711.537	***
	(848.51)		(660.96)		(1575.82)	
Formal Self-employed to Formal Self-employed	12097.630	***	10537.072	***	12434.895	***
	(1192.00)		(1703.91)		(2141.84)	
Formal Self-employed to Formal Worker	5827.793	***	5278.981	***	2437.215	**
	(1076.09)		(1523.93)		(1051.56)	
Formal Worker to Unemployed	-1576.436	***	-1940.113	***	-2161.025	***
	(328.03)		(237.81)		(427.38)	
Formal Worker to Informal Worker	3037.783	***	1848.114	***	2779.085	***
	(257.88)		(152.20)		(275.47)	
Formal Worker to Informal Self-employed	4406.599	***	2244.568	***	3726.094	***
	(313.88)		(192.18)		(340.96)	

Table 4.3 (Continued)

Formal Worker to Formal Self-employed	8718.413	***	7992.312	***	9208.727	***
	(1282.34)		(1268.39)		(1670.31)	
Formal Worker to Formal Worker	3558.109	***	2541.005	***	3418.881	***
	(253.50)		(160.34)		(287.57)	
Region						
Mexico City (omitted)						
US Border	548.806	***	608.806	***	727.502	***
	(84.88)		(79.31)		(117.82)	
North	85.365		109.581		276.262	***
	(76.47)		(67.14)		(91.18)	
Center	-79.721		-191.689	***	-18.502	
	(57.41)		(48.92)		(73.89)	
South	-464.985	***	-361.778	***	-467.909	***
	(85.22)		(67.79)		(109.58)	
Constant	-6545.706	***	-3419.910	***	-3433.571	***
	(645.21)		(516.08)		(687.17)	
R-squared	0.333		0.376		0.331	
N	98327		83112		55415	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: OLS Regression by Periods. Logarithms.
Dep. Var.: Change in Reported Earnings

	Q1:87-Q2:94		Q3:94-Q1:99		Q2:99-Q4:02	
	Coef./(s.e.)		Coef./(s.e.)		Coef./(s.e.)	
Initial Log-earnings	-0.532	***	-0.510	***	-0.512	***
	(0.01)		(0.01)		(0.01)	
Age						
Linear	0.025	***	0.023	***	0.015	***
	(0.00)		(0.00)		(0.00)	
Squared	-0.000	***	-0.000	***	-0.000	***
	(0.00)		(0.00)		(0.00)	
Education						
Linear	0.007	***	0.003		-0.009	***
	(0.00)		(0.00)		(0.00)	
Squared	0.001	***	0.002	***	0.002	***
	(0.00)		(0.00)		(0.00)	
Male	0.148	***	0.152	***	0.156	***
	(0.01)		(0.01)		(0.01)	
Sector Transitions						
Informal Worker to Informal Self-employed	0.205	***	0.173	***	0.148	***
	(0.03)		(0.03)		(0.03)	
Informal Worker to Formal Self-employed	0.818	***	0.985	***	0.847	***
	(0.20)		(0.22)		(0.13)	
Informal Worker to Formal Worker	0.140	***	0.163	***	0.155	***
	(0.02)		(0.02)		(0.02)	
Informal Self-employed to Informal Worker	-0.064	**	-0.125	***	-0.148	***
	(0.03)		(0.03)		(0.03)	
Informal Self-employed to Informal Self-employed	0.130	***	0.105	***	0.075	***
	(0.02)		(0.02)		(0.02)	
Informal Self-employed to Formal Self-employed	0.583	***	0.667	***	0.628	***
	(0.06)		(0.15)		(0.06)	
Informal Self-employed to Formal Worker	0.090	***	0.145	***	0.133	***
	(0.03)		(0.02)		(0.03)	
Formal Self-employed to Informal Worker	-0.010		-0.222		-0.617	
	(0.13)		(0.17)		(0.44)	
Formal Self-employed to Informal Self-employed	0.259	***	0.231	***	0.222	**
	(0.05)		(0.05)		(0.10)	
Formal Self-employed to Formal Self-employed	0.665	***	0.739	***	0.631	***
	(0.05)		(0.06)		(0.08)	
Formal Self-employed to Formal Worker	0.281	***	0.342	***	0.054	
	(0.06)		(0.09)		(0.09)	
Formal Worker to Informal Worker	-0.001		-0.016		-0.015	
	(0.02)		(0.02)		(0.03)	
Formal Worker to Informal Self-employed	0.192	***	0.065	**	0.140	***
	(0.03)		(0.03)		(0.03)	
Formal Worker to Formal Self-employed	0.532	***	0.674	***	0.649	***
	(0.08)		(0.09)		(0.12)	
Formal Worker to Formal Worker	0.140	***	0.219	***	0.166	***
	(0.02)		(0.01)		(0.01)	
Region						
Mexico City (omitted)						
US Border	0.119	***	0.167	***	0.146	***
	(0.01)		(0.01)		(0.01)	
North	0.018	**	0.020	**	0.064	***
	(0.01)		(0.01)		(0.01)	
Center	0.014	*	-0.018	**	0.016	*
	(0.01)		(0.01)		(0.01)	
South	-0.060	***	-0.084	***	-0.104	***
	(0.01)		(0.01)		(0.01)	
Constant	3.397	***	3.068	***	3.470	***
	(0.08)		(0.07)		(0.09)	
R-squared	0.269		0.286		0.279	
N	95607		79650		53854	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

4.4.3 Measurement Error

As previously mentioned in the methodological section, a simulation is performed in order to appraise the potential effects of measurement error on the unconditional mobility estimates.

The simulation based on equation (4.9), i.e.,

$$\hat{\beta}_1 = \beta_1 \frac{V(y^*_{it-1})}{V(y_{it-1})} + (\rho - 1)\frac{\alpha(2+\alpha)V(\varepsilon_{it-1})}{V(y_{it-1})} + (\theta - 1)\frac{V(\zeta_{it-1})}{V(y_{it-1})}$$

consists in assuming that the true $\beta_1 = 0$, i.e., the mobility profiles are unrelated to initial earnings, and it is asked "How big should the measurement error be in order to lead to the conclusion of convergence, when in fact there is none?". More specifically, the simulation tries to capture how big should the variance for the measurement error component be as a fraction of the total variance of *reported* earnings, in order to obtain convergence of the magnitudes observed.

Under the assumption of no convergence in true earnings, i.e if $\beta_1 = 0$, there are 4 unknowns in equation (4.9): the α parameter that arises because of the correlation between measurement error and true earnings, the autocorrelation parameter ρ in transitory earnings, the variance of this term $V(\varepsilon_{it-1})$, and the autocorrelation parameter θ in the idiosyncratic component of measurement error.[24] Since these are too many parameters to identify, the simulation here presented will further assume $\theta = 0$ and equal variances between transitory earnings and the idiosyncratic component of measurement error, i.e., $V(\varepsilon_{it-1}) = V(\zeta_{it-1})$. Assuming $\theta = 0$ just makes stronger the potential impact of measurement error. This is because, for a given variance of the measurement error, higher values of θ would make $\hat{\beta}_1$ bigger, making it harder to find convergence as a result of measurement error. The

[24]The variance of $V(\zeta_{it-1})$ can be obtained from the total variance in reported earnings if α and $V(\varepsilon_{it-1})$ are known.

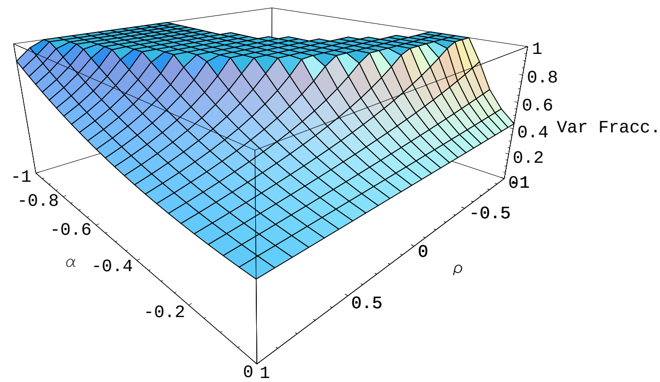


Figure 4.15: Measurement error Simulation

assumption that transitory earnings and the idiosyncratic measurement error have equal variances has no further basis than pure convenience.[25] Finally, the value of the parameter $\hat{\beta}_1$ selected for the simulation is -0.438, which is the average of the $\hat{\beta}_{1t}$ calculated for each panel t with Least Squares.

Figure 4.15 shows the ratio of the variance of the measurement error to the variance of initial reported earnings, as a function of ρ and α. This ratio is sometimes called the *noise-to-signal* ratio. The graph plots how big this ratio must be in order to give an OLS parameter of $\hat{\beta}_{1t} = -0.438$. Hence, the lower this ratio is on the graph, the more pernicious is the effect of measurement error on the OLS estimates, (i.e. it is easier for them to be biased towards finding convergence).

This graph shows that in order for the convergence result to be completely due to measurement error, the noise-to-signal ratio needs to be *at least* 40% (when $\alpha = 0$), and higher if α is negative. This is a relatively high noise-to-signal ratio. For comparison purposes, for the U.S. Bound and Krueger (1991) found this

[25]In the study of Pischke (1995) for the U.S., the magnitudes of these components are found to be roughly the same, but this result varied for different time periods.

ratio to be around 28% in their sample of men in the CPS. Although, without further information coming from validation studies applied to Mexico, it cannot be evaluated whether such numbers are too high or not, it seems unlikely that the convergence result obtained is entirely due to measurement error.[26]

The second issue studied is the impact of measurement error on the parameter estimates of the conditional mobility equation (4.4). In equation (4.10), i.e.

$$\Delta y_{it} = (\rho_t - 1)y_{it-1} + Z_i\tilde{\gamma}_t + \sum_l \sum_m st(l,m)\pi_{lm} + ((\alpha+1)\eta_{it} + (\theta-\rho)\zeta_{it-1} + \omega_{it})$$

it was shown that if earnings were measured with error, initial reported earnings would be correlated with the error term. This, in turn would lead to biased estimates of all the parameters in the equation, if estimated by Least Squares. The approach taken here is to use a series of regressors as instruments for the initial level of earnings y_{it-1}, and use the predicted measure of $\hat{y}_{it-1}$ as a regressor in the conditional mobility equation. The instruments include all the regressors in (4.4) (i.e., age, education, gender, sector, and regional dummies), as well as wealth proxies and occupation dummies.

Instead of doing a simple 2-Stages Least Squares (2SLS) to correct for the measurement error bias, I estimate a whole system of equations by 3-Stages Least Squares (3SLS). The equations jointly estimated are

$$\begin{aligned}
y_{it-1} &= Z_i\gamma_{t-1} + X_{it-1}\kappa_{t-1} + (\alpha\varepsilon_{it-1} + \zeta_{it-1}) \\
y_{it} &= Z_i\gamma_t + X_{it}\kappa_t + (\alpha\varepsilon_{it} + \zeta_{it}) \\
\Delta y_{it} &= (\rho_t - 1)y_{it-1} + Z_i\tilde{\gamma}_t + \sum_l \sum_m st(l,m)\pi_{lm} + ((\alpha+1)\eta_{it} + (\theta-\rho)\zeta_{it-1} + \omega_{it})
\end{aligned}$$

[26]There is a region of values for α, ρ where it is impossible to generate such convergence pattern. In particular, the range of values forming the flat part on the top of Figure 4.15, and the area that follows afterwards, are ranges of values where convergence cannot arise.

Estimating this system by 3SLS partially corrects the measurement error bias (in the same way the standard 2SLS-IV estimator does), but in addition to that it can be used to perform a specification test for the structure of the earnings model described in equations (4.2)-(4.4). The assumed structural form is testable, since equation (4.4) (the third equation on the system) was derived from the first two equations by assuming that the error term ε_{it} is autocorrelated. In particular, the restrictions $\tilde{\gamma}_t = \gamma_t - \rho_t \gamma_{t-1}$ and $\pi_{lm} = \kappa_t(m) - \rho_t \kappa_{t-1}(l)$ are testable using the information from earnings equations in the first and final periods, together with the mobility equation. Such test provides information on whether the assumed structure is rejected by the data or not.

The results of the 3SLS estimations are presented in two parts. First, the parameter estimates for $(\hat{\rho}_t^{3SLS} - 1)$ are presented in Figure 4.16. Then the full regression results for the pooled subperiods are included in Tables 4.5 and 4.6.

As it can be appreciated in Figure 4.16 once the instrumentation is performed the conditional mobility appears to be divergent for some of the early periods in the sample, but it becomes convergent afterwards. In general, the estimated conditional convergence rates are smaller than the ones estimated via LS.[27]

The analysis of the full regression results in Tables 4.5 and 4.6 shows some differences with respect to the OLS estimations in Tables 4.3 and 4.4. In particular, the effects of human capital variables on mobility are smaller. Age variables are not always statistically significant, exhibiting a increasing concave pattern only during the aftermath of the Peso crisis. Education has a positive effect most of the periods in the specification in levels, but after the Peso crisis the U-pattern appears again with an inflexion point around 6 yrs. of education. This convex pattern is

[27]The standard errors in the 3SLS estimations are quite small most likely due to the use of extra information coming from the two earnings equations.

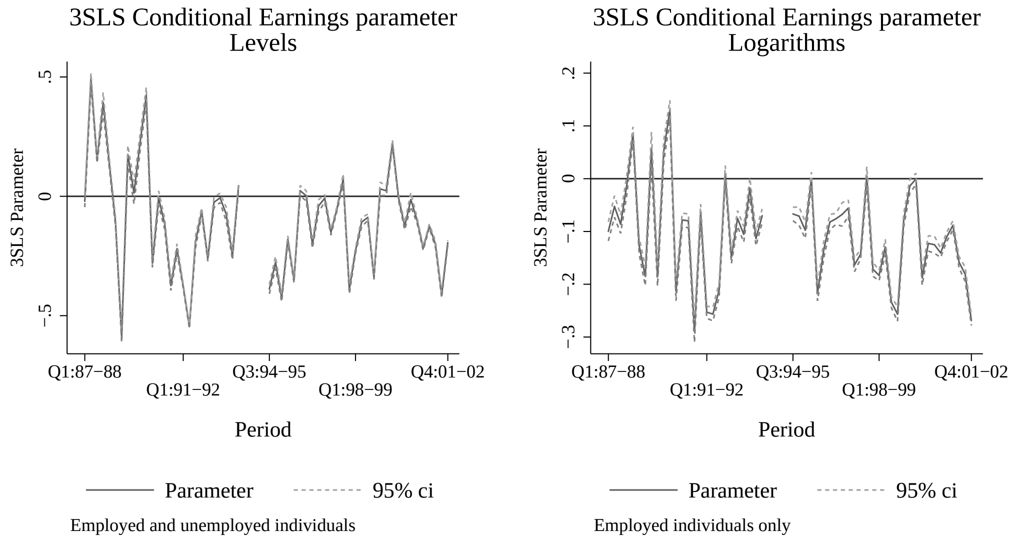


Figure 4.16: 3SLS. Conditional Mobility Parameter

also found in most of the periods for the logarithmic specification, with an inflexion point fluctuating between 6 and 10 years of education. Being male has a positive effect on mobility, but the effect is much smaller than the one obtained in the LS regression. Finally, the patterns of conditional mobility by sector transition look similar to the ones estimated under LS. The regional dummies are not significant in many cases, but for most of the cases the US Border and North regions have a higher conditional mobility.[28]

[28]With the exception of the period 1987-93 for the US Border cities.

Table 4.5: 3SLS Regression by Periods. Levels. Dep. Var.: Change in Reported Earnings

	Q1:87-Q2:94	Q3:94-Q1:99	Q2:99-Q4:02
	Coef./(s.e.)	Coef./(s.e.)	Coef./(s.e.)
Initial Earnings	-0.043 ***	-0.187 ***	-0.096 ***
	(0.00)	(0.00)	(0.00)
Age			
Linear	11.147	34.551 **	-33.537 *
	(17.58)	(14.19)	(19.73)
Squared	-0.194	-0.450 **	0.346
	(0.22)	(0.18)	(0.24)
Education			
Linear	-17.619	-59.777 ***	-22.235
	(14.98)	(12.46)	(17.85)
Squared	2.301 ***	5.395 ***	2.127 **
	(0.74)	(0.59)	(0.84)
Male	95.379 **	70.771 **	62.262
	(40.28)	(31.68)	(43.60)
Sector Transitions			
Unemployed to Informal Worker	4332.743 ***	3217.986 ***	4296.387 ***
	(155.43)	(103.84)	(173.67)
Unemployed to Informal Self-employed	5649.411 ***	3976.265 ***	5285.783 ***
	(146.82)	(98.56)	(168.86)
Unemployed to Formal Self-employed	13180.098 ***	12411.062 ***	13286.223 ***
	(482.10)	(292.52)	(574.57)
Unemployed to Formal Worker	4888.759 ***	4085.576 ***	5154.720 ***
	(138.09)	(91.93)	(159.33)
Informal Worker to Unemployed	-3827.637 ***	-2653.410 ***	-3655.622 ***
	(156.54)	(99.18)	(172.06)
Informal Worker to Informal Worker	493.490 ***	578.438 ***	644.917 ***
	(188.54)	(116.38)	(208.29)
Informal Worker to Informal Self-employed	1801.668 ***	1324.286 ***	1579.085 ***
	(186.54)	(116.46)	(207.45)
Informal Worker to Formal Self-employed	9111.391 ***	9050.538 ***	9616.414 ***
	(279.79)	(385.59)	(367.22)
Informal Worker to Formal Worker	1042.026 ***	1394.146 ***	1458.778 ***
	(184.53)	(114.53)	(205.23)
Informal Self-employed to Unemployed	-4979.779 ***	-3368.817 ***	-4493.590 ***
	(147.09)	(93.00)	(163.61)
Informal Self-employed to Informal Worker	-635.633 ***	-130.672	-194.015
	(186.48)	(116.42)	(207.69)
Informal Self-employed to Informal Self-employed	670.471 ***	618.488 ***	745.816 ***
	(180.38)	(110.97)	(202.38)
Informal Self-employed to Formal Self-employed	8015.395 ***	8294.042 ***	8752.673 ***
	(227.60)	(178.64)	(265.68)
Informal Self-employed to Formal Worker	-95.897	688.850 ***	629.868 ***
	(180.35)	(112.10)	(202.41)
Formal Self-employed to Unemployed	-1.10e+04 ***	-9626.783 ***	-1.16e+04 ***
	(280.26)	(201.61)	(614.66)
Formal Self-employed to Informal Worker	-6473.768 ***	-6303.426 ***	-7283.754 ***
	(257.47)	(248.90)	(328.12)
Formal Self-employed to Informal Self-employed	-5136.017 ***	-5551.103 ***	-6399.542 ***
	(218.29)	(163.02)	(261.32)
Formal Self-employed to Formal Self-employed	2126.333 ***	2167.275 ***	1667.414 ***
	(247.91)	(196.46)	(295.24)
Formal Self-employed to Formal Worker	-5920.108 ***	-5383.258 ***	-6532.291 ***
	(219.83)	(166.09)	(261.98)
Formal Worker to Unemployed	-4272.175 ***	-3337.792 ***	-4375.583 ***
	(137.25)	(83.74)	(148.90)
Formal Worker to Informal Worker	102.979	-36.649	-11.295
	(184.54)	(115.08)	(206.24)
Formal Worker to Informal Self-employed	1408.169 ***	695.048 ***	924.396 ***
	(180.26)	(112.31)	(202.97)

Table 4.5 (Continued)

Formal Worker to Formal Self-employed	8685.365	***	8402.423	***	8915.297	***
	(229.01)		(186.66)		(267.00)	
Formal Worker to Formal Worker	646.397	***	784.346	***	802.634	***
	(176.80)		(107.71)		(199.05)	
Region						
Mexico City (omitted)						
US Border	-178.441	**	233.129	***	133.992	*
	(80.22)		(56.90)		(77.18)	
North	53.027		75.213	**	136.869	***
	(48.37)		(37.50)		(52.03)	
Center	20.800		-31.795		127.379	**
	(49.21)		(40.61)		(55.71)	
South	-92.753		-137.895		-54.463	
	(125.75)		(91.28)		(117.07)	
Constant	-594.602		-942.625	***	384.103	
	(387.20)		(298.34)		(438.81)	
R-squared	0.0534		0.2086		0.1139	
N	98318		83107		55411	

* p < 0.1, ** p < 0.05, *** p < 0.01

Table 4.6: 3SLS Regression by Periods. Dep. Var.: Change in Reported Log-Earnings

	Q1:87-Q2:94		Q3:94-Q1:99		Q2:99-Q4:02	
	Coef./(s.e.)		Coef./(s.e.)		Coef./(s.e.)	
Initial Log-Earnings	-0.0330	***	-0.1057	***	-0.0894	***
	(0.001)		(0.001)		(0.002)	
Age						
Linear	-0.0005		0.0040	**	-0.0029	
	(0.002)		(0.002)		(0.002)	
Squared	-0.0000		-0.0001	**	0.0000	
	(0.000)		(0.000)		(0.000)	
Education						
Linear	-0.0051	***	-0.0018		-0.0081	***
	(0.002)		(0.002)		(0.002)	
Squared	0.0004	***	0.0004	***	0.0004	***
	(0.000)		(0.000)		(0.000)	
Male	-0.0143	***	0.0210	***	0.0111	**
	(0.004)		(0.004)		(0.005)	
Sector Transitions						
Informal Worker to Informal Worker (omitted)						
Informal Worker to Informal Self-employed	0.2033	***	0.1760	***	0.1689	***
	(0.007)		(0.007)		(0.008)	
Informal Worker to Formal Self-employed	0.7596	***	0.9149	***	0.8380	***
	(0.022)		(0.040)		(0.035)	
Informal Worker to Formal Worker	0.1843	***	0.3125	***	0.2602	***
	(0.006)		(0.006)		(0.007)	
Informal Self-employed to Informal Worker	-0.1915	***	-0.1580	***	-0.1732	***
	(0.007)		(0.006)		(0.008)	
Informal Self-employed to Informal Self-employed	0.0102		0.0181	**	-0.0035	
	(0.009)		(0.008)		(0.010)	
Informal Self-employed to Formal Self-employed	0.5637	***	0.7513	***	0.6593	***
	(0.018)		(0.020)		(0.023)	
Informal Self-employed to Formal Worker	-0.0077		0.1571	***	0.0891	***
	(0.009)		(0.008)		(0.010)	
Formal Self-employed to Informal Worker	-0.6897	***	-0.7362	***	-0.7897	***
	(0.019)		(0.026)		(0.030)	
Formal Self-employed to Informal Self-employed	-0.4881	***	-0.5537	***	-0.6136	***
	(0.016)		(0.018)		(0.022)	
Formal Self-employed to Formal Self-employed	0.0635	***	0.1782	***	0.0488	*
	(0.020)		(0.023)		(0.027)	
Formal Self-employed to Formal Worker	-0.5067	***	-0.4145	***	-0.5251	***
	(0.016)		(0.018)		(0.022)	
Formal Worker to Informal Worker	-0.1796	***	-0.2439	***	-0.2317	***
	(0.006)		(0.006)		(0.007)	
Formal Worker to Informal Self-employed	0.0219	**	-0.0714	***	-0.0638	***
	(0.009)		(0.008)		(0.010)	
Formal Worker to Formal Self-employed	0.5733	***	0.6602	***	0.5989	***
	(0.018)		(0.021)		(0.023)	
Formal Worker to Formal Worker	0.0037		0.0683	***	0.0276	***
	(0.008)		(0.007)		(0.009)	
Region						
Mexico City (omitted)						
US Border	-0.0295	***	0.0553	***	0.0174	*
	(0.009)		(0.008)		(0.009)	
North	0.0003		0.0201	***	0.0341	***
	(0.005)		(0.005)		(0.006)	
Center	0.0026		-0.0020		0.0298	***
	(0.005)		(0.006)		(0.007)	
South	-0.0135		-0.0165		-0.0071	
	(0.013)		(0.012)		(0.014)	
Constant	0.3332	***	0.6335	***	0.8357	***
	(0.039)		(0.040)		(0.048)	
R-squared	0.0332		0.1050		0.0933	
N	95602		79649		53851	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Specification Tests for the Conditional Earnings Model

	Levels		Logarithms	
Period	χ stat	p-value	χ stat	p-value
Q1:87-Q2:93	25.11	0.456	6.06	0.993
Q3:94-Q1:99	59.21	0.000	96.99	0.000
Q2:99-Q4:01	13.29	0.973	13.45	0.706

To end this section the results for the tests of the structural relationships are presented in Table 4.7. The test is a joint test of the hypotheses $\tilde{\gamma}_t = \gamma_t - \rho_t \gamma_{t-1}$ and $\pi_{lm} = \kappa_t(m) - \rho_t \kappa_{t-1}(l)$. The results show that the data fail to reject these hypotheses during the first and last periods of the sample. This is a positive thing, since it means that the assumed structure does not contradict the data. However, for the years in the aftermath of the Peso crisis the data strongly rejects the model proposed. This means that the shock that came after this crisis altered the dynamic structure of earnings, and mobility cannot be described by the same model that explained reasonably well the other periods. A more careful study on what is the structure of earnings dynamics during this period is an interesting topic that deserves further research.

4.4.4 Attrition Bias and Non-Response

In this section the potential effects of attrition and non-reporting in the panel are studied. In particular, the bounds

$$E(\Delta y_{it} | y_{it-1}, z_i = 1) P(z_i = 1 | y_{it-1})$$

$$E(\Delta y_{it} | y_{it-1}, z_i = 1) P(z_i = 1 | y_{it-1}) + P(z_i = 0 | y_{it-1})$$

of partial identification of $E(\Delta y_{it}|y_{it-1})$ are presented.

Figure 4.17 shows the estimated kernel and the partial identification bounds for the unconditional mobility expectation $E(\Delta y_{it}|y_{it-1})$. The graph plots the results for some select years. The solid line in the figures displays the kernel estimates for the individuals with complete information. As expected from the previous analysis, these estimates have a negative slope, especially for high-earners. The dotted lines in the graph denote the lower and upper bounds of the identification region $H[E(\Delta y_{it}|y_{it-1})]$. Recalling, the meaning of this region is that, without further assumptions, $E(\Delta y_{it}|y_{it-1})$ can lie anywhere inside these bounds.

As it can be seen, for 1987 and 1989 it cannot be ruled out that the mobility was not convergent, since the identification region contains the zero line inside it.[29] For 1995 and 2001 the result of convergence still holds in the presence of attrition. In general, the results change depending on the year selected, but the loss of information due to attrition and non-reporting is substantial. One interesting result to note is that the bounds become wider for rich individuals. The reason for this is that rich individuals are more prone to not reporting their earnings, hence the probability of including them in the sample is smaller.

The results presented only focused on the impact of attrition in the unconditional mobility estimates, but they are indicative of the perverse effects of this problem in terms of loss of information. This loss of information will be carried to other mobility estimates (e.g. conditional mobility parameters, aggregate mobility indices, etc.) using this data.

Before closing this section it is important to remark that the bounds previously presented are the more negative scenario that can be faced. In particular, unlike

[29]Notice however, that the bounds are much narrower for 1987. This is because attrition was considerably lower in that year.

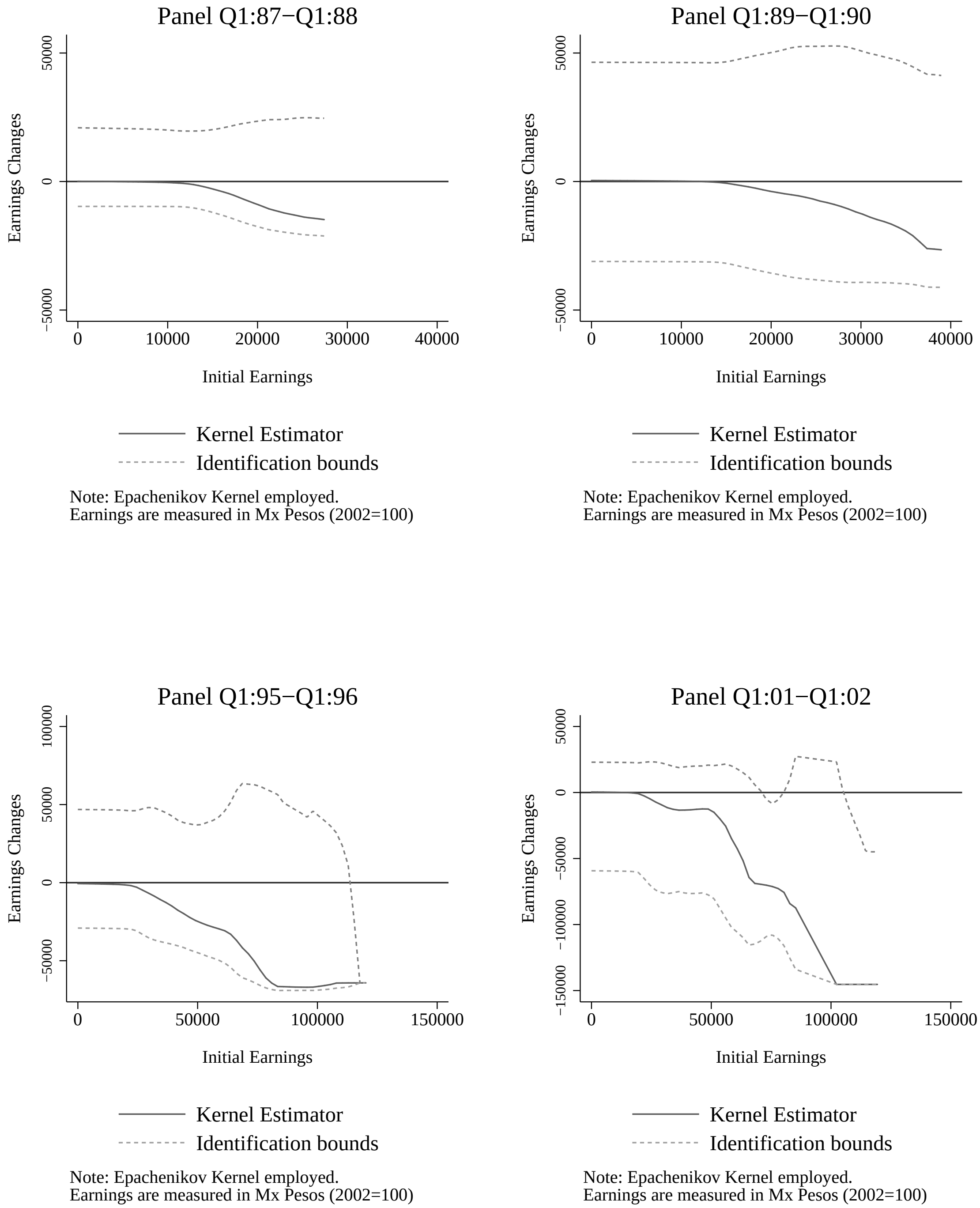


Figure 4.17: Partial Identification Bounds on Unconditional Mobility Expectation

other treatments of the attrition problem, no assumptions about the attrition process were made. If one starts building up extra assumptions (e.g., a fraction x of the population with missing values is actually missing at random), tighter bounds can be obtained and more positive conclusions will be reached. One interesting line of research to pursue would be to find a good instrument that is related to earnings, but not to attrition.[30]

This section, rather than giving a completely negative panorama of the effects of missing data on the analysis, attempts to call for caution in the presence of such high levels of attrition. The results presented throughout the dissertation are still meaningful for the subsample of the population with complete information. To what extent they can be extended to the whole urban population is something that requires more analysis.

4.5 Conclusions

This chapter studied the relationship between different measures of initial advantage and earnings mobility, as well as the impact of socioeconomic characteristics of the individual on earnings mobility.

The answer to the question of "Are the most advantaged individuals gaining more (losing less) in terms of earnings changes?" is that no; in the majority of the estimations the most advantaged individuals either keep their advantage or lose more than the rest. When reported earnings are taken as the measure of initial advantage, having a higher initial advantage is found to be negatively associated

[30]Clearly, such an instrument is difficult to find since, if earnings affect the probability of exiting the panel or of not reporting earnings, then any variable causing earnings can not be used as an instrument. The candidates must be variables caused by earnings or having a common cause with earnings.

with mobility. In other words, there is convergence in earnings between high-earners and low-earners. This convergence fluctuates over time, but its magnitude is rather stable. However, there is evidence suggesting that this result is reflecting an adjustment of earnings from a transitory shock back to its more permanent level. For this reason, when the relationship between mobility and several proxy measures for permanent advantage is considered, there is much less convergence or no convergence at all. In these estimations, a low convergence pattern is found during the late nineties-early 00's, and the only case of strong convergence in earnings occurs right after the 1994 Peso crisis. This episode led to convergence in earnings levels, but not in logarithms. This indicates that during this period high-earners lost more than everybody else in absolute terms, but their losses were proportional to their higher initial earnings.

In general, what these results imply is that, most of the times, the mobility experienced over a year does not alter the permanent advantage of individuals. These results should lead to reinterpreting some of the conclusions reached in Chapter 3. In particular, the conclusion that mobility equalized earnings over time between initial earnings quintile groups is most likely arising because of adjustments of earnings to transitory shocks. As shown in this chapter, it was only during the aftermath of the 1994 Peso crisis that mobility led to convergence between the earnings of permanently advantaged individuals and the ones of individuals permanently disadvantaged. In these years, the negative effects of the crisis were spread throughout the economy and not even high-earners managed to avoid losses.

Regarding the answers to the question "What is the *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility?" in general, it is found that *ceteris paribus* more education leads to negative mobility for individuals with low levels of education, but after a certain point more education is associated with upward conditional mobility (the inflexion point fluctuates between 3 and 10 years of education depending on the model). Being male has a strong positive effect on mobility, and age has a small, but positive effect also.

Holding everything else constant, transitions into formal self-employment are the ones that bring the largest gains (smallest losses), while transitions into informal wage work are associated with the largest losses (smallest gains), excluding of course the movements into unemployment. Transitions into formal wage work and informal self-employment are between the two previous categories. In general, it seems that transitions to informal self-employment bring more positive mobility than the ones into formal wage work. However, this result is sometimes reversed in the period following the 1994 Peso crisis. One problem with the previous results is that in the case of self-employed individuals, it is not possible to discern how much of the reported earnings are payments to the labor factor, and how much are returns to physical capital. Finally, living in cities along the US Border and in the North brings upward conditional mobility, while living in the Center and South brings more negative conditional mobility.

Finally, the answer to the question of "How does the impact of initial earnings on mobility change once the effect of these determinants is accounted for?" it is found that, holding everything else constant, initial earnings are negatively related to earnings mobility, meaning that individuals converge to their own conditional mean. This conditional convergence rate is homogenous for most of the population, except for the cases of education and sector groups. In these cases, the individuals with higher earnings (the ones with higher education and the formal self-employed, respectively) present smaller convergence rates, i.e., for these individuals a shock to their earnings was more persistent. Also, the conditional convergence rate is slightly stronger than the unconditional one. This means that the overall impact of the individual socioeconomic characteristics generates divergence in earnings.

Simulations on the impact of measurement error show that this error needs to be quite large in order to be the *sole* reason underlying the convergence results found in the unconditional mobility regression with reported earnings. Needless to say, not much else can be said without a proper validation study on the amount of measurement error that applies for a country like Mexico. It is important to remark that the results found for the unconditional relationship between mobility and permanent advantage are not affected by this measurement error. The effects of measurement error on the conditional mobility estimates cannot be fully controlled, but the instrumented conditional mobility estimates are smaller than the ones obtained under the assumption of no measurement error.

The amount of attrition in the panel and of non-reporting of the earnings variable leads to great losses of information that can create serious problems to the estimations previously presented. In this chapter, the effects of these two problems are assessed by means of partial identification analysis. This analysis estimates a whole region in which the conditional expectations of interest may lie, given the amount of information revealed by the data.

The partial identification bounds estimated are large and, depending on the year selected, they sometimes challenge the convergence results obtained in the unconditional mobility estimations. Although the case analyzed in this chapter is the worst case scenario in terms of the potential impact of attrition, the findings obtained warn against generalizing the results to the whole urban population without further research on this topic.

4.6 Appendix

4.6.1 Proofs for Expressions in Section 4.3.4

This appendix presents derivations of equation (4.9) and establishes the consistency of the IV parameter estimating the relationship between earnings mobility and permanent advantage.

Proof of (4.9)

Note first that, under the measurement error model described by (4.5)-(4.8), the estimated covariance between reported mobility and initial reported earnings equals

$$cov(\Delta y_{it}, y_{it-1}) = cov(\Delta y_{it}^P + (1+\alpha)\Delta\varepsilon_{it} + \Delta\zeta_{it}, y_{it-1}^P + (1+\alpha)\varepsilon_{it-1} + \zeta_{it-1})$$

with $y_{it}^P \perp \varepsilon_{it} \perp \zeta_{it}$. Hence

$$\begin{aligned} cov(\Delta y_{it}, y_{it-1}) &= cov(\Delta y_{it}^P, y_{it-1}^P) + (1+\alpha)^2 cov(\Delta\varepsilon_{it}, \varepsilon_{it-1}) + cov(\Delta\zeta_{it}, \zeta_{it-1}) \\ &= cov(\Delta y_{it}^*, y_{it-1}^*) + \alpha(2+\alpha)(\rho-1)V(\varepsilon_{it-1}) + (\theta-1)V(\zeta_{it-1}) \end{aligned}$$

Dividing this expression by $V(y_{it-1})$ and recalling that the true (unbiased) $\beta_1 = cov(\Delta y_{it}^*, y_{it-1}^*)/V(y_{it-1}^*)$, gives the biased OLS $\hat{\beta}_1$ parameter in expression (4.9).

Proof of unbiasedness of the IV parameter estimating the relationship between earnings mobility and predicted permanent advantage.

First notice that the unbiasedness of $\hat{y}_{it-1}$ follows because if measured earnings y_{it} equal $y_{it} = y_{it}^P + (1+\alpha)\varepsilon_{it} + \zeta_{it}$, and if permanent earnings are the component of earnings determined by a vector of variables W_{it} affecting permanent advantage, i.e., $y_{it}^P = W_{it}\Gamma_t$, then the first-stage regression of y_{it-1} on W_{it} will give unbiased estimators of Γ_t, by the assumed orthogonality between y_{it}^P, ε_{it}, and ζ_{it}. With these $\hat{\Gamma}_t$ the predicted $\hat{y}_{it-1}$ will also be unbiased.

The unbiasedness of $\hat{y}_{it-1}$ being established, it easily follows that the second-stage regression

$$\Delta y_{it} = \beta_0^P + \beta_1^P \hat{y}_{it-1} + u_{it}$$

will give unbiased estimators of β_1^P, by the aforementioned orthogonality conditions and the fact that $\Delta y_{it} = \Delta y_{it}^P + (1+\alpha)\Delta\varepsilon_{it} + \Delta\zeta_{it}$.

4.6.2 Convergence Parameter Estimates With and Without the Unemployed

Figure 4.18 shows that when estimating equation (4.1) in levels it does not make a difference whether unemployed individuals are included or not. In both cases, the estimated parameters are barely distinguishable from each other.

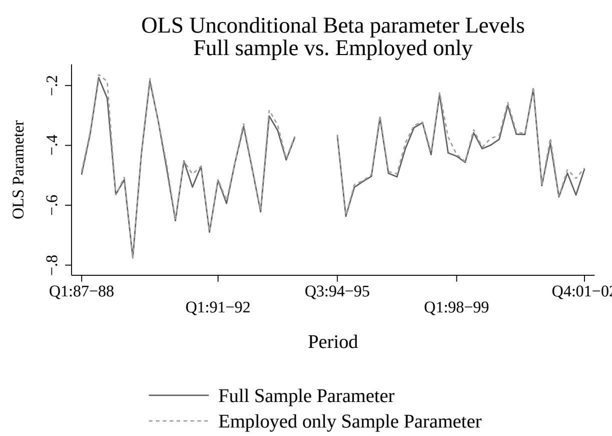


Figure 4.18: OLS Unconditional Mobility Parameter. With and Without Unemployed.

Chapter 5

Testing Segmentation in Mexican Labor Markets

5.1 Introduction

In the previous chapter it was shown that sector transitions played an important role in affecting earnings mobility. The results of chapters 3 and 4 show that, both unconditionally and conditionally, there is a large premium for being formal when comparing formal wage work to informal wage work, and formal self-employment to informal self-employment as destination sectors. Whether this constitutes evidence for segmentation or not, is unclear. The possibility of endogeneity between earnings mobility and the choice of sector makes the interpretation of the previous comparisons inconclusive.

This chapter studies the issue of segmentation more closely, in particular it tests whether individuals working in the informal sector have free access to formal sector jobs, or whether formal sector jobs are rationed. It also analyzes the implications of the structure of labor markets for earnings mobility.

The theory of labor market segmentation has a long tradition in economics that dates back to the works of Furnivall (1939) and Boeke (1953), but more importantly to the seminal work of Lewis (1954). This theory stipulates that the functioning of the labor markets is not fully competitive. Instead, it posits that two or more markets can operate at the same time, and in these markets individuals of equal productive capacity get paid differently. More importantly, it states that individuals cannot freely move into jobs in the better paying sectors because such jobs are rationed. Although the original dualistic models were mainly concerned with explaining the disparities between the urban industrialized areas versus the backward rural areas in developing economies, soon extensions of this model incorporated the analysis of urban informal markets (see for instance Fields, 1975). These models considered the informal sector as a second-best option that workers take when they cannot find a formal sector job in the city.

This chapter tries to test whether such a model of segmentation applies for the case of urban Mexico, and if that is the case, what are the earnings losses of those individuals rationed out of the formal sector. It also tries to assess how many of these segmented individuals manage to enter the formal sector in subsequent periods, and what are the earnings gains associated with such transitions.

5.2 Previous Literature and Contribution

The literature on labor market segmentation has a long tradition in modern economics. Its origins can be traced back to the seminal work of Lewis (1954) on Dualism. This theory tried to explain the existence of an advanced urban industrialized sector coexisting with a backward rural area in developing countries. This model was quickly extended in many directions, many of which are not directly related to the topic here studied.[1]

One of these extensions however, was the source for the study of segmentation in urban labor markets. The seminal work of Harris and Todaro (1970) explored the implications of this dualistic structure on rural-urban migration. In particular, it provided an explanation for the migration flows from rural areas into the better

[1]For recent appraisals of this dualistic approach see Basu (1997); Fields (2004b) and Ranis (forthcoming).

paid urban industrialized ones, even in the presence of urban unemployment. The main logic of this model is that jobs are rationed in the industrialized urban sector and, assuming downward wage rigidities in the urban areas, rural migration would take place up to the point where the rural marginal product of labor equals the expected wage in the urban sector (the expectation coming from the fact that the probability of finding an urban job is less than one).

The Harris-Todaro model was soon extended in many ways, one of which incorporated the existence of an urban informal sector. The model proposed by Fields (1975) explained the creation of an urban informal sector as the result of the need of urban unemployed workers to have some source of income. In other words, many workers in urban areas couldn't afford being unemployed and would set up (or work at) an informal business that would provide them with a source of income while waiting to gain access to a better paid formal sector job. This characterization of the informal sector as a buffer zone, that provided second-class types of jobs for workers rationed out of the formal sector, was very popular for many years in the Development Economics literature. However, in the last decades it has has been questioned.

Another tradition in the literature that proposed the existence of segmented labor markets is the one started by the work of Doeringer and Piore (1971), which had as main tenets the existence of low-wage jobs, usually with low returns to human capital and bad working conditions, together with the rationing of good jobs. This literature, contrary to the literature started by Lewis, arose as a way to explain labor markets in *developed* economies, and generated on its own a vast amount of research (see Cain, 1976; Dickens and Lang, 1985, for a review of the literature).

Almost parallel to the development of this literature another tradition in economics viewed the existence of several sectors in the labor market as the product of the natural alignment of comparative advantage forces. The seminal paper of Roy (1951) provided a rationale for sectoral allocation according to the comparative advantages that each individual had at different economic activities. The model also showed how this sector allocation could generate a skewed distribution of income. Soon, the framework set by the Roy model became the workhorse for studying not only sectoral allocation issues, but many others like occupational choice, migration decisions, unionization, marital status, to mention just a few. A recent appraisal of this model exploring its empirical content can be found on Heckman and Honoré (1990).

A crucial feature of the Roy model relevant for the present discussion is that it assumed that individuals are free to move across sectors. In other words, although the Roy model can explain the co-existence of "good jobs" and "bad jobs", it differs from the dualistic models in that it assumes no rationing among sectors for this to occur. Under this framework, if the informal sector is a worse paid/bad conditions sector it is because of its inherent technological characteristics, and the productive characteristics of the workers that self-select themselves to work there.

These two competing views of the informal labor markets have crucial differences in terms of what they would recommend as policies to ameliorate the conditions of informal sector workers. In its most simple form, a dualistic model of labor markets would call for an expansion of the formal sector so that it could absorb the workers stuck in the informal sector. On the other hand, if the informal sector arose because of the self-selection mechanism posited by the Roy model, expanding the formal sector *per se* would not solve the problem. Instead a better option would be to enhance the productivity of the informal sector workers.

Since these models present contrasting views on how the labor markets operate, it is important to empirically test which one of them (if any) provides a better explanation of the labor markets. This is particularly relevant in developing economies, due to the large size of the informal sector in these economies.

There have been several attempts to test for the segmentation of labor markets. Although initially it was thought that the poor quality of informal sector jobs was enough evidence to prove that formal sector jobs were rationed, economists soon realized that a formal way of testing this assumption was necessary. In the beginning, tests were attempted by classifying workers into good and bad jobs according to their job characteristics, and comparing either the earnings or the earnings functions of workers with comparable human capital across these two types of jobs. This approach was abandoned because if a Roy model was true and workers could freely move across sectors, and if this decision depended on unmeasured factors, then these comparisons would not test segmentation, but rather just document that the formal and the informal sectors were in fact different (Rosenzweig, 1988, provides a discussion of these and other criticisms).

Among the first papers that explicitly tried to account for issues of self-selection while testing for segmentation was Dickens and Lang (1985). Their model was a switching regression model with unknown separation, i.e., it allowed for individuals to self-select into sectors, but it did not assume beforehand the sector the workers were in. Instead, this feature was implicitly estimated in the model. For this model, evidence of segmentation would be if a two-wage equation specification, with one wage profile increasing in human capital variables and the other one being flat, is preferred over a single equation model. In their application using PSID data for

1980, the authors find evidence of segmentation in US labor markets.

This type of tests came under criticisms of several kinds. In a paper published shortly after, Heckman and Hotz (1986) questioned the validity of such a test. The most important criticism directed at this methodology was that the evidence favoring a two-wage equation specification could indicate instead the existence of a highly nonlinear wage equation.[2]

Another test of segmentation was proposed by Magnac (1991) in an application to Colombian women in the 80's. His model was a multivariate Tobit used to analyze the joint decision between not entering the labor force, entering the informal sector, entering the formal sector, or being unemployed (if no job is found). The test of segmentation was a test of relative costs of entry into the formal sector. These costs were assumed to be proportional to wages, and could be interpreted as queueing into the formal sector. Magnac found no evidence of segmentation with this data; however, specification tests of his model turned out to give little support to the distributional assumptions implicit in the model.[3]

A series of papers have tried to test segmentation by comparing wage equations corrected for self-selection into the formal and informal sectors, but the methods used to correct for sample selectivity assume for starters that individuals choose their sector of employment in order to maximize either their income or their utility. This assumption is problematic, since this free choice among sectors is precisely the issue being tested. Papers following this approach are Gindling (1991) for Costa

[2] A recent application of a methodology similar to the one of Dickens and Lang, but addressing some of the criticisms of Heckman and Hotz, is Basch and Paredes-Molina (1996) for Chile. The authors claim to find evidence of dualism, but not of rationing.

[3] In particular, the assumptions of quadrivariate normality and homoskedasticity were rejected by the data.

Rica, Marcouiller *et al.* (1997) for Mexico, Peru and El Salvador, and Thomas and Vallée (1996) for Cameroon. Since their methodology rules out from the beginning the possibility of job rationing in the formal sector, it is hard to know how to interpret the diverse results presented in those papers.

An interesting approach has been proposed by Pradhan and van Soest (1995) in their study of segmentation in urban Bolivia in the late eighties. In this paper the authors study whether an Ordered Probit (which assumes a ranked choice among Formal sector jobs, Informal sector jobs, and Unemployment) describes the data better than a Multinomial Probit, which assumes no ranking among the options. The results they find is that the Multinomial Probit is a better model for women, while for men no model can be preferred. Although this model is an interesting empirical application, it has the disadvantage of not modeling explicitly the rationing process.

Two recent approaches to test for segmentation are the ones proposed by Pisani and Pagán (2003) and Navarro-Lozano and Schrimpf (2004). The first paper provides a test of segmentation by explicitly modeling the rationing of formal sector jobs, while the second one generates a full distribution of counterfactual wages (and not only expected values, as usually done in the literature) for workers in the formal and informal sector. Since this dissertation follows the methodology proposed by Pisani and Pagán (2003), the methodology of this paper will be discussed more carefully later on. Regarding the second paper, it will be discussed in the section dealing with applications of this literature to the Mexican case.

Finally, a recent test based on propensity score matching has been proposed by Pratap and Quintin (forthcoming). In this paper the authors compare the wages of formal sector workers with a control wage generated by a weighted average of wages of informal sector workers, matched by the propensity score method. The authors also perform Difference-in-Difference estimations to control for the effects of unobserved heterogeneity. Segmentation is considered to exist in this model if a formal sector wage premium is found. The authors apply their method to urban Argentina in the early nineties and they find no evidence of segmentation. Although this method has the advantage of not relying on distributional assumptions as the previous ones, it considers monetary factors only. Excluding non-monetary factors could lead to wrong inferences; consider for instance if formal sector non-monetary benefits outweigh by far the benefits of working in the informal sector, then workers could be willing to accept a pay cut in the formal sector in order to receive those benefits.[4]

5.2.1 Studies on Labor Market Segmentation for Mexico

Out of the topics studied in this dissertation, informality and segmentation in labor markets is by far the one that has received more attention for the Mexican case. In a long series of papers, William Maloney has questioned the segmented view of the labor markets for Mexico (and Latin America) (Maloney, 1998, 1999, 2003; Maloney *et al.*, 2004; Maloney, 2004; Maloney and Bosch, 2005, are some examples of this literature). In a nutshell, Maloney questions the idea that informal self-employment is an undesirable state where workers end only when they cannot find formal sector jobs. Instead, he proposes that many of the informal self-employed are willingly so, because the overall utility they derive from these jobs is higher than the one they would obtain in the formal sector. He argues that many times formal

[4]Actually in their empirical application the authors consider further tests of job satisfaction, but due to data limitations the evidence they present is not conclusive.

sector benefits are of low quality; and that, in the case of Mexico, informal workers can enjoy such benefits if a family member is employed in the formal sector. Also, informal self-employment allows workers to have flexible working hours, something that the overregulated formal labor market does not allow. This flexibility in hours is particularly important for female workers. Finally, he notes that many of these informal self-employed have low human capital and hence would receive a low pay in the formal sector if they were to enter it.

The fact that the informal self-employment sector is a heterogeneous one, having individuals with different reasons for being there has been acknowledged before by proponents of the segmented view of the labor markets (see for instance Fields, 1990). However, recognizing this heterogeneity in the informal sector assumes the existence of a group of workers who are restricted from obtaining a formal sector job. Furthermore, most of the evidence in the aforementioned papers has been concerned with the informal self-employed, and much less has been said about the informal wage workers (who work for the owner of an informal firm). While the studies of Maloney have gathered a fair amount of evidence supporting this heterogeneity in the informal sector, the issue of how many of the informal workers are informal by force and not by choice, cannot be resolved unless a direct test of segmentation is performed.

A recent paper that tries to directly test segmentation for the Mexican case is Navarro-Lozano and Schrimpf (2004). In this paper, the authors propose a Monte Carlo Markov Chain method similar to the one used in Carneiro *et al.* (2003), where a joint model of earnings, sector choice and schooling is estimated. The model is comprised of a set of mutually independent factors that jointly affect the aforementioned dependent variables, and that account for the unmeasured characteristics of the individuals. Under this framework, segmentation would occur if (after controlling for a set of individual characteristics) workers in the informal sector experience (on average) gains by moving to the formal sector that are higher than the average gains of workers already there. This constitutes a test of segmentation because it would mean that, *ceteris paribus*, the movers started at a lower position; something that these workers could have avoided if sector mobility was unconstrained.[5] The authors test their results with the ENEU in 1997, and exclude the self-employed from their estimations. They find no evidence of segmentation for this year. While the model proposed by Navarro-Lozano and Schrimpf is richer than any other previously encountered in the literature, it has a serious drawback. The problem is that the way the authors model sector choice, implicitly assumes that workers are free to move across sectors in order to maximize their utility. This, as previously mentioned, is problematic because this assumption is precisely what is being tested.

Finally, a set of studies that, although do not test directly for segmentation, they provide interesting evidence on sector transitions and earnings mobility is Gong *et al.* (2000) and Gong and van Soest (2002). The first paper uses a dynamic Multinomial Logit to analyze the transition patterns between the informal and formal sector, and non-participation. The authors use the ENEU for periods pre and post the 1994 peso crisis. Their main findings are that entry and exit rates for the formal sector are lower than for the informal sector, the probability of formal sector employment increases with education level, while the probability of working

[5]This comparison of average gains (rather than earnings levels), conditional on a set of individual characteristics, allows for testing segmentation even if the agents are utility maximizers (instead of just income maximizers). This occurs if one assumes, as the authors do, that the difference in non-monetary benefits between sectors is fixed over time.

in the informal sector decreases with family income. The authors interpret these results as providing support for the view that the informal sector is a temporary state for workers rationed out of the formal sector.

The second paper (Gong and van Soest, 2002) analyzes jointly quarterly wage mobility and sector transitions (between formal and informal jobs) by means of a dynamic Random effects model. Their main findings in the wage equation are that wage differentials between the formal and the informal sector increase with education level, and that the lagged sector state does not affect current wages. In the sector transition equation they find that the probability of formal sector employment strongly increases with the wage differential between formal and informal sector jobs. Although this paper is one of the first attempts to jointly model earnings and sector choice in a dynamic setting for Mexico, it does not incorporates the possibility of rationing in the sector choice equation.

The contribution of the present chapter is then to use the segmentation test proposed in Pisani and Pagán (2003), apply it to the Mexican case, and extend their application by estimating selectivity-corrected earnings equations. The estimation of these earnings equations is crucial to obtain counterfactual earnings for the individuals restricted out of the formal sector. These predicted earnings serve to estimate the potential gains that such individuals could experience by moving into the formal sector. In addition to that, this chapter exploits the panel structure of the data to provide evidence on sectoral transitions and the *actual* earnings mobility experienced by individuals who were predicted to be restricted from entering the formal sector in the initial period.

5.3 Methodology

5.3.1 Sector Choice

In order to test for segmentation in this chapter the workers' decision to apply for a formal sector job is modeled together with the employer's decision to accept or not such applicants. This sector choice model is the same of Pisani and Pagán (2003), which in turn base their analysis on the models of Poirier (1980) and Abowd and Farber (1982).

The propensity of a worker i for working in the formal sector (denoted by V_i^w) is assumed to be a linear function (in the parameters) of a set of socioeconomic characteristics Z, i.e.,

$$V_i^w = Z_i \gamma_w + v_{1i} \tag{5.1}$$

Among the factors that may influence the worker's desire for a formal sector job are his age, education level, gender, the region in which he works, his wealth, his marital status, the family structure in his household, his desire to be an entrepreneur instead of a salaried worker, his ability to perform well in a formal sector job, among many other things. Not all of these factors are indeed observed by the econometrician. In the present application, the variables observed that are included in the empirical specification are human capital variables (age and education), gender, regional dummies, proxies for household wealth, and family structure variables (a dichotomous variable for marital status and the number of children and adults in the household). Unobserved factors like ability and desire for entrepreneurship will be captured by the v_{1i} term.

It is important to note that some of these factors may operate through many channels. For instance, the workers' human capital variables (like age and education) may affect the desire for a formal sector job through their effect on earnings, but they may also have a direct effect on V_i^w. Note then that the specification (5.1) is a reduced form that does not distinguish among these many channels.

Similarly, for the formal sector employer, a set of the worker characteristics X, as well as his own characteristics will determine whether he offers or not a job to the applicant. In particular, it will be assumed that the propensity for hiring an applicant (denoted by V_i^e) will be given by

$$V_i^e = X_i\gamma_e + v_{2i} \tag{5.2}$$

In this case it is assumed that the variables included in X are a subset of Z, i.e., $X \subset Z$. This means that when deciding whether to hire or not an applicant a formal sector employer only takes into account those characteristics that affect the productivity and performance of the worker in the firm. In particular the variables that constitute the X vector are human capital variables (age and education), gender, and regional dummies for controls. Other characteristics that affect the decision of the worker for applying to formal sector jobs, but that do not affect his productivity in the firm (like household structure and wealth proxies), do not influence the decision of the employer. An additional justification for why these variables do not enter X is that, even if they were to have an effect on productivity, they are not observed by the employer. As it will be clear below, these exclusion restrictions imposed on X are crucial to the identification of the main components of the model.

In the specification (5.2) it is also assumed the characteristics of the employer will not be explicitly included as observed variables, instead they will be captured in v_{2i}, for reasons that will be explained below.

With equations (5.1) and (5.2) it is possible to specify the probability of a worker applying to a formal sector job, the probability of being hired conditional on having applied to such job, and finally the probability of being a formal worker. More specifically,

$$P(Apply = 1|Z_i) = P(V_i^w > 0|Z_i) = P(v_{1i} > -Z_i\gamma_w)$$

$$P(Hire = 1|X_i, Apply = 1) = P(V_i^e > 0|X_i, Apply = 1) = P(v_{2i} > -X_i\gamma_e)$$

$$P(FormalWorker = 1|Z_i) = P(Apply = 1|Z_i)P(Hire = 1|X_i, Apply = 1)$$

The problem with this structure is that in practice one does not observe whether an individual applied for a formal sector job or not. The ENEU in particular asks unemployed individuals whether they had tried to start their own business or searched for salaried work, but there is no indication about the whether those options would take place in the formal or the informal sector. For the employed individuals the question is even less informative since they only ask them whether they have looked for another job.

In order to identify the separate probabilities, and answer whether there is segmentation in labor markets, a bivariate probit model with partial observability is applied. The origin of these models is Poirier (1980), where the conditions for identification and estimation of the parameters of interest are stated. The version applied in this chapter is closer to the simpler version of this model proposed by Abowd and Farber (1982) in their study of queueing and union status in the US. In particular, this chapter assumes that the error terms (v_{1i}, v_{2i}) in (5.1) and (5.2) are bivariate normally distributed, i.e., $(v_{1i}, v_{2i}) \sim N(0, \Sigma)$, but the correlation between these terms is zero, i.e., the errors are independent. Although this is a restrictive assumption,[6] it will be used in this first approach, because the full model that allows for a non-zero correlation between these error terms is harder to esti-

[6]It makes sense to believe that there are unobservable factors to the econome-

mate.[7] As Maddala (1983) notes, assuming independence between the error terms (v_{1i}, v_{2i}) is equivalent to assuming that the sector choice process is sequential, i.e., first the worker applies to a formal sector job, and then a formal sector employer decides whether to hire or not the applicant. A task for future work is to extend the present methodology allowing for a non-zero correlation between these error terms, and hence opening the possibility for a joint decision process.

The likelihood function for this model is

$$L = \prod_{FW=1} [\Phi(Z_i\gamma_w)\Phi(X_i\gamma_e)] \prod_{FW=0} [1 - \Phi(Z_i\gamma_w)\Phi(X_i\gamma_e)] \tag{5.3}$$

where FW is the dichotomous indicator of whether an individual is a formal wage worker or not, and $\Phi(\cdot)$ is the standard normal distribution function.

Note that under this model the parameters γ_w, γ_e for (5.1) and (5.2) are estimated from a single dependent variable (the indicator FW). This is why it is crucial to have exclusion restrictions from the set of variables that affect the employer's decision. Without these exclusion restrictions on X it would be impossible to separately identify $\Phi(Z_i\gamma_w)$ from $\Phi(X_i\gamma_e)$ in (5.3).

A difference between the likelihood model used in this chapter and the one used by Abowd and Farber (1982) is that they use extra information on the job rights of workers to assume that certain workers are not restricted from entering the union. This restriction turns out to give them strong identifying power in their application, but in the present context there is no variable that could play an equivalent role.[8] In any case the present application does not appear to need

trician that both the applicant and the employer observe and act upon. Consider for instance the punctuality with which the worker showed up at the job interview.

[7]In such versions it is hard to obtain convergence of the Maximum Likelihood algorithm, and many times the solutions lie on the boundary of the parameter space, preventing the computation of the variances.

[8]An attempt was made to estimate a model similar to theirs by assuming that

such extra identifying assumptions in order to reach a solution.

From this structure it is also clear why the few employer characteristics collected by the survey are not included as observable variables in (5.1) and (5.2). Since the survey contains information about the employer characteristics (mainly sector of economic activity and firm size) only for the job at which the individual is currently employed, it is not clear how to introduce them in the model. Consider for instance how to interpret if one were to find that a dummy for working at service activities impacts negatively $P(Hire = 1|X_i, Apply = 1)$. Does it means that formal sector jobs in services are harder to get? Or is it that individuals that got rejected in other types of formal sector jobs actually turned to providing services in the informal sector? What one would like to include is the characteristics of the potential employers that the individual is considering, and not the firm variables after the selection process has taken place.

This bivariate probit model nests as a particular case the standard probit which assumes no segmentation. If the formal sector jobs are not rationed, then all that would matter is whether a worker wants to be in the formal sector or not. Under this structure a standard univariate probit applied over (5.1) would estimate all the relevant parameters in the model.[9] This nesting gives the basis for testing whether formal sector jobs are rationed or not. In particular, a Likelihood Ratio (LR) test

---

any worker who is in the formal sector in the second period, who was also a formal sector worker in the previous period and did not change sector of economic activity (meaning manufacturing, services, construction, etc.) nor occupation, was unrestricted from entering the formal sector, i.e., for this workers $P(Hire = 1|X_i, Apply = 1) = 1$. This version of the model failed to converge to a solution and hence it was abandoned.

[9] To see that this is indeed a particular case of the more general bivariate model, note that the univariate model can be derived from the bivariate likelihood function in equation (5.3) if $\gamma_e = 0$ for all the individual variables and the constant term γ_{0e} is set high enough as to make $\Phi(X_i\gamma_e) = 1$.

can be applied to see whether the bivariate model describes the data better than the univariate probit model. If the LR test favors the richer structure (5.1) and (5.2) over the single probit model, *and* if there is a group of individuals for whom the model predicts rejection in the formal sector, i.e., $P(Hire = 0|X_i, Apply = 1)$, then this will constitute evidence of segmentation in labor markets.[10]

Applying this methodology to test for segmentation in the labor markets was first done by Pisani and Pagán (2003). They made an application to the Nicaraguan labor markets and found evidence of segmentation. The present chapter extends their application by estimating earnings equations for each sector, correcting for selectivity bias. The methodology for such estimations is presented in the next section.

Before moving on to the next section it is important to mention that the analysis performed excludes formal self-employed individuals[11] and unemployed individuals. The reason for excluding the unemployed workers is because the bivariate probit structure proposed does not have room for the third choice of waiting another period to keep searching for a better job. Furthermore, it is not evident how to discern which of the unemployed are waiting for a better offer, and which ones just cannot find any job at all. In any case, the number of unemployed individuals in the subsample under study (individuals between 25-60 yrs. of age in the labor force) is extremely small and hence their exclusion is likely to make no difference for the estimations.

In the case of the formal self-employed, their exclusion from the sample is

[10]It is important to notice that this model could underestimate the degree of segmentation if there are individuals who are discouraged from looking for formal sector jobs.

[11]That is individuals who declare to be self-employed or bosses in firms that are registered with the government or provide some sort of coverage to their workers.

done because it is not clear what are the restrictions a worker faces in order to become a formal self-employed. While in the case of a formal wage worker it is clear that in principle an applicant would go through a screening process by the employer; in the case of a formal self-employed the role played by the access to capital necessary to start a formal sector firm seems to be crucial. Since the dataset used in this dissertation is not rich on measures of capital, it is better to exclude such individuals rather than wrongly impute to them a sector choice structure.

Also since this criticism on the importance of access to capital might extend to a fraction of the informal self-employed, all the estimations in this chapter are performed with *and* without these informal self-employed, to test if the conclusions change.

5.3.2 Earnings Equations

An extension of the present chapter to the work of Pisani and Pagán (2003) is to move beyond the estimation of the discrete process (5.1) and (5.2), and estimate earnings equations for each sector corrected for selectivity bias. This allows obtaining the effects of socioeconomic variables on earnings free of sample selectivity bias. These estimates also allow estimating counterfactual earnings in the formal sector for individuals working in the informal sector.

The earnings of the individual are related to his productive characteristics X by

$$y_{iI} = X_i\beta + \varepsilon_i \tag{5.4}$$

if he works in the informal sector, and by

$$y_{iF} = X_i(\beta + \pi) + \varepsilon_i \tag{5.5}$$

if he works in the formal sector.[12]

In practice the earnings of an individual are observed in only one sector at a time, and the individuals are not randomly assigned into sectors. Hence, estimating equations (5.4) and (5.5) by standard regression methods would lead to biased estimates. In order to obtain unbiased estimates of the parameters in (5.4) and (5.5), an adjustment for selectivity bias is required.

The selectivity-correction model followed in this chapter assumes trivariate normality between the error terms in equations (5.1), (5.2), and (5.5), i.e., $(v_{1i}, v_{2i}, \varepsilon_i) \sim N(0, \Sigma)$.[13]

Under these assumptions, it can be shown (see the appendix of this chapter) that the expectation of the error term ε_i, conditional on the individual being in the formal sector, is

$$E(\varepsilon_i | FW = 1) = \rho_{wF}\lambda_{wF} + \rho_{eF}\lambda_{eF}$$

with[14]

$$\lambda_{wF} = \frac{\phi(Z_i\gamma_w)}{\Phi(Z_i\gamma_w)} \qquad \lambda_{eF} = \frac{\phi(X_i\gamma_e)}{\Phi(X_i\gamma_e)} \tag{5.6}$$

and the expectation of ε_i conditional on the individual being in the informal sector is

$$E(\varepsilon_i | FW = 0) = \rho_{wI}\lambda_{wI} + \rho_{eI}\lambda_{eI}$$

[12]The assumption that earnings in different sectors differ by a parametric shift π, but keep the same error structure ε_i, is necessary to obtain selectivity-corrected estimates for the earnings equation in the informal sector. For an application with similar assumptions see Tunalı (1986).

[13]This is equivalent to assuming that the error terms (v_{1i}, v_{2i}) are bivariate normal as previously stated in equations (5.1)-(5.2), and that the expectation of ε_i conditional on (v_{1i}, v_{2i}) and Z_i is linear. The independence between v_{1i} and v_{2i} is maintained.

[14]$\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal density and distribution functions respectively.

with

$$\lambda_{wI} = \frac{-\phi(Z_i\gamma_w)\Phi(X_i\gamma_e)}{1-\Phi(Z_i\gamma_w)\Phi(X_i\gamma_e)} \qquad \lambda_{eI} = \frac{-\phi(X_i\gamma_e)\Phi(Z_i\gamma_w)}{1-\Phi(Z_i\gamma_w)\Phi(X_i\gamma_e)}. \tag{5.7}$$

The terms λ_{wF}, λ_{eF}, λ_{wI}, and λ_{eI} are the equivalent to the familiar Inverse Mills' Ratio in the standard Heckman selectivity correction model. In this case, there are two correction terms for each sector, because in addition to the worker's decision to apply for a formal sector job, there is the decision of formal sector employer on whether to hire or not a given applicant.

With this structure, the selectivity-corrected earnings equations to estimate are

$$y_{iI} = X_i\beta + \rho_{wI}\lambda_{wI} + \rho_{eI}\lambda_{eI} + \varepsilon_i \tag{5.8}$$

$$y_{iF} = X_i(\beta + \pi) + \rho_{wF}\lambda_{wF} + \rho_{eF}\lambda_{eF} + \varepsilon_i \tag{5.9}$$

The λ-terms can be constructed by using the estimates of γ_w and γ_e that arise from eqn. (5.3). This two-step method has been analyzed in Poirier (1980) and Tunalı (1986). Since these papers do not include the derivation of the selectivity correction terms for the case when the individual is in the informal sector, the derivation of these expressions in its more general form is included in the appendix of this chapter. Expressions (5.6) and (5.7) are particular cases of that derivation.

Once equations (5.8) and (5.9) are estimated this permits generating counterfactual expected earnings for a random individual with characteristics X, in the sector in which he is not participating. These counterfactual predictions y_i^F and y_i^I can be compared to the actual earnings of the individual. In particular, these counterfactuals provide an estimate of the potential monetary gains of moving into the formal sector for individuals restricted from entering that sector.

It is important to note that both the discrete choice models (5.1)-(5.2) as well as the earnings equations (5.8) and (5.9), are not dynamic. The generalization of this type of models to a dynamic setting has not been done yet in the literature. For this reason these models are estimated at the first period at which individuals are interviewed. Using the first interview for each individual also eliminates any potential negative effect of attrition from the panel. For the same compactness of presentation reasons adduced in previous chapters, the models are estimated for the pooled periods going from the 1st quarter 1987 to the 2nd quarter of 1993, the 3rd quarter of 1994 to the 1st quarter of 1999, and the 2nd quarter of 1999 to the 4th quarter of 2002.

Although the full panel is not exploited in the estimation of the structural models, its information is used in exploring the evolution of earnings and sector transitions for the individuals for whom the model was originally estimated. The two main questions analyzed there will be: i) "How many of the individuals predicted to be restricted from entering the formal sector in period 1 enter into that sector in subsequent periods?" and ii) "What is the *actual* earnings mobility experienced by rationed individuals if they manage to enter the formal sector in further periods?". The last question is formulated both unconditionally and controlling for a set of individual characteristics. Analyzing these two issues sheds some light on how easy it is to gain access to the formal sector in subsequent periods, and what are the *actual* gains (as opposed to the predicted ones) of the individuals that managed to enter the formal sector.

5.4 Results

5.4.1 Sector Choice

The first results presented in this section pertain to the probit analysis for the sector choice process. The results of these estimations are shown in Tables 5.1 - 5.4. The estimations are performed for two different specifications, one including informal self-employed, and the other keeping only wage workers in the sample. In each of these tables two sets of estimates are presented, one that assumes free entry to the formal wage jobs (denominated "No Rationing Model"), and one that allows for formal sector jobs to be rationed (labeled "Rationing Model").

At the bottom of each of these tables the Likelihood Ratio statistic is reported for the hypothesis test that the free entry model describes better the data. As it can be seen, this hypothesis is soundly rejected in all of the specifications. In other words, under the stated distributional and functional assumptions, this test provides the basis for the claim that urban labor markets in Mexico are segmented. This result is of relevance because it goes against all the previous findings reported in the literature applied to Mexico.[15] This means that if the implicit assumptions of the empirical model are correct, in urban Mexico it would be incorrect to treat entry into the formal sector as depending only on the decision of the worker. Instead, whenever dealing with issues of sector choice, an explicit account for the possibility of formal sector jobs rationing should be allowed.

The analysis of the "No Rationing" models, that assume free entry into the formal sector, shows that, holding everything else constant, being younger, more educated and female affects positively the probability of being formal in the model

[15] This literature was reviewed in section 5.2 of the chapter.

that includes self-employed workers (Tables 5.1 and 5.2). However, when the observations are restricted to only wage workers the model indicates that being older, more educated and male affects positively the probability of being a formal salaried worker (Tables 5.3 and 5.4).

The regional effects vary depending on the specification, and on the periods under scrutiny. In the model with self-employed, being in the North affects positively the probability of being formal, while being on the US border affects it negatively in the period going from 1987 to 1993, but positively after 1994. The model with wage workers only, shows strong positive effects of being in a region other than Mexico City.

Being married is sometimes positively associated with being a formal sector worker, and having small children in the household is always negatively associated with it. The effect of the number of adults living in the household varies depending on the specification. In the model including self-employed it is always positive, but not so in the model with wage workers only.

Finally, the effects of wealth proxy variables, like cluster average income and dwelling characteristics are positive on the probability of being formal.[16]

Moving now to the analysis of the model where rationing is allowed, it is important to remember that here there are two equations being estimated. Namely, the application decision of the worker to a formal sector job (reported under the heading "Pr(A=1)" in the tables), and the formal sector employer's decision to accept such applicant (reported under the heading "Pr(E=1| A=1)").

[16]The only exception is the variable "Type of Dwelling". This is a categorical variable that is *decreasing* in the quality of the type of dwelling the household lives in. For instance, it takes the value 1 if the household lives in an owned house, 2 if in an apartment, 4 if in a room on a roof, etc.

Table 5.1: Probit of Sector Allocation. Informal Self-employed Included.Q1:87-Q2:93

	No Rationing Model		Rationing Model			
	Pr(Formal=1)		*Pr(A=1)*		*Pr(E=1|A=1)*	
Age						
Linear	-0.0125	***	-0.0053		-0.0094	***
	(0.005)		(0.011)		(0.003)	
Squared	0.0000		-0.0003	**	0.0000	
	(0.000)		(0.000)		(0.000)	
Education						
Linear	0.1261	***	0.0231	**	0.0852	***
	(0.004)		(0.012)		(0.003)	
Squared	-0.0029	***	0.0139	***	-0.0017	***
	(0.000)		(0.001)		(0.000)	
Male	-0.0525	***	3.0431	***	-0.4723	***
	(0.012)		(0.455)		(0.013)	
Region						
Mexico City (omitted)						
US Border	-0.0464	***	0.8199	***	-0.1451	***
	(0.013)		(0.089)		(0.014)	
North	0.0543	***	-0.2632	***	0.0776	***
	(0.011)		(0.031)		(0.010)	
Center	-0.0931	***	-0.1888	***	-0.0929	***
	(0.011)		(0.029)		(0.009)	
South	-0.0221		0.0021		-0.0385	
	(0.017)		(0.087)		(0.023)	
Married	0.0298	***	-0.4688	***		
	(0.013)		(0.025)			
Household Structure						
# hh members >11yrs	0.0143	***	0.0733	***		
	(0.003)		(0.006)			
# hh members <12yrs	-0.0313	***	-0.0993	***		
	(0.004)		(0.008)			
Wealth Proxies						
Cluster Average Income	0.0000		0.0001	***		
	(0.000)		(0.000)			
Constant	0.0956		0.3238		0.6679	***
	(0.098)		(0.229)		(0.065)	
No. Observations	197,221				197,221	
Log-Likelihood	-117,547.8				-115,708.6	
LRT statistic	3678.4		p-value		0.000	

* p < 0.1, ** p < 0.05, *** p < 0.01

The decision of applying for a formal sector job is positively affected by the education of the worker. The effect of age varies depending on the model and the period considered. Most of the times being older increases the probability of applying for a formal sector job, although sometimes no effect is found.

Table 5.2: Probit Models of Sector Allocation. Informal Self-employed included. Q3:94-Q4:01.

	Q3:94-Q1:99			*Q2:99-Q4:01*		
	No Rationing Model	Rationing Model		No Rationing Model	Rationing Model	
	Pr(Formal=1)	*Pr(A=1)*	*Pr(E=1|A=1)*	*Pr(Formal=1)*	*Pr(A=1)*	*Pr(E=1|A=1)*
Age						
Linear	-0.0255 ***	-0.0002	-0.0226 ***	-0.0167 **	0.0438 **	-0.0205 ***
	(0.006)	(0.015)	(0.004)	(0.007)	(0.017)	(0.005)
Squared	0.0001 **	-0.0002	0.0001 ***	0.0001	-0.0008 ***	0.0001 **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Linear	0.1002 ***	0.0349 *	0.0699 ***	0.0990 ***	-0.0746 ***	0.0773 ***
	(0.005)	(0.018)	(0.005)	(0.006)	(0.022)	(0.005)
Squared	-0.0017 ***	0.0077 ***	-0.0008 ***	-0.0014 ***	0.0160 ***	-0.0010 ***
	(0.000)	(0.002)	(0.000)	(0.000)	(0.002)	(0.000)
Male	-0.0684 ***	1.7939 ***	-0.4017 ***	-0.0893 ***	1.8218 ***	-0.3755 ***
	(0.013)	(0.120)	(0.015)	(0.015)	(0.145)	(0.016)
Region						
Mexico City (omitted)						
US Border	0.3044 ***	1.3216 ***	0.1053 ***	0.3554 ***	2.1843 ***	0.1596 ***
	(0.019)	(0.133)	(0.017)	(0.023)	(0.720)	(0.019)
North	0.1214 ***	-0.0853 **	0.1477 ***	0.2077 ***	0.1172 *	0.2012 ***
	(0.015)	(0.043)	(0.012)	(0.018)	(0.061)	(0.014)
Center	-0.0719 ***	-0.1733 ***	-0.0356 ***	0.0367 **	-0.1963 ***	0.0815 ***
	(0.015)	(0.044)	(0.011)	(0.018)	(0.053)	(0.014)
South	0.1076 ***	-0.2335 ***	0.1603 ***	0.1789 ***	-0.1031	0.2194 ***
	(0.022)	(0.089)	(0.032)	(0.027)	(0.107)	(0.036)
Married	0.0135	-0.4213 ***		0.0266	-0.4831 ***	
	(0.015)	(0.033)		(0.018)	(0.043)	
Household Structure						
# hh members >11yrs	0.0124 ***	0.0617 ***		0.0165 ***	0.1109 ***	
	(0.004)	(0.008)		(0.005)	(0.013)	
# hh members <12yrs	-0.0223 ***	-0.0881 ***		-0.0198 ***	-0.0906 ***	
	(0.005)	(0.011)		(0.007)	(0.016)	
Wealth Proxies						
Cluster Average Income	0.0000	0.0001 ***		0.0000	0.0001 ***	
	(0.000)	(0.000)		(0.000)	(0.000)	
Type of Dwelling	0.0660 ***	0.1852 ***		0.0706 ***	0.0865 ***	
	(0.011)	(0.023)		(0.013)	(0.027)	
Solid Walls	0.0308	0.0408		0.0016	0.0259	
	(0.025)	(0.064)		(0.033)	(0.095)	
Solid Roofs	0.0951 ***	0.1956 ***		0.0959 ***	0.1787 ***	
	(0.022)	(0.039)		(0.027)	(0.051)	

Table 5.2 (Continued)

Solid Floors	0.0764	***	0.2776	***			0.0688	***	0.1779	***		
	(0.015)		(0.034)				(0.018)		(0.041)			
Electricity	0.1592	**	0.3942	**			0.1264		0.5793	**		
	(0.079)		(0.194)				(0.125)		(0.265)			
Water	-0.0077		-0.0573				0.1167	*	0.1149			
	(0.039)		(0.083)				(0.061)		(0.118)			
Sewage	0.0694	**	0.0873				0.0332		0.1594			
	(0.033)		(0.069)				(0.041)		(0.106)			
Phone	0.0158		0.1216	***			-0.0183		0.0459			
	(0.015)		(0.033)				(0.017)		(0.039)			
Other services	-0.0202		-0.1097	**			0.0493	***	0.1160	*		
	(0.017)		(0.047)				(0.019)		(0.061)			
Kitchen	0.1179	***	0.2280	***			0.0648	***	0.1140	**		
	(0.020)		(0.035)				(0.024)		(0.045)			
Constant	-0.2938	**	-0.9647	***	0.7320	***	-0.5392	***	-1.8320	***	0.5569	***
	(0.139)		(0.364)		(0.085)		(0.186)		(0.466)		(0.099)	
No. Observations	139,685				139,685		97,570				97,570	
Log-Likelihood	-85,908.7				-85,234.6		-59,700.2				-59,192.7	
LRT statistic	1348.2		p-value		0.000		1015.0		p-value		0.000	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Being a male has a large positive effect on the probability of applying to a formal sector job in the model including self-employed, but it has a negative effect in the model with wage workers only. This would mean that for a woman it is more attractive to look for informal self-employment, probably because women, being the persons doing all the housework, prefer jobs that allow them more flexible schedules. Being married has a negative effect on the decision to apply for a formal sector job in the sample including the self-employed, but it has a positive effect in the sample with wage workers only.

The number of children in the household diminishes the probability of applying for a formal sector job under all specifications (somebody has to take care of the children and formal sector jobs do not offer flexible schedules). However, while the number of adults increases the probability of applying to such jobs in the sample with informal self-employed, the effect becomes negative in the sample with wage workers only. This is a puzzling finding because independently of the sector the worker is in, having fewer adults around who can take care of the children should lead to a smaller desire for jobs that are not flexible in schedules.[17]

One variable that was not included in the analysis, but that might play a role in the decisions analyzed, is whether the individual has a relative working in the formal sector. This could discourage applying to formal sector jobs, because in Mexico individuals can get health coverage if a close relative has a formal sector job. The reason for not incorporating this variable in the analysis is that doing so would require moving away from the framework of individual sector participation,

[17]While this finding could be due to an income effect, by which more adults in the household lead to greater income per-capita, and hence allow the individual worker to look for less demanding jobs in the informal sector, this does not explain why the effect changes as we move from one sample to the other.

to one of household labor supply, where joint decisions are taken on "who applies for jobs, where". This is a more complicated model and won't be pursued here.

The impact of regional variables on the decision of applying for a formal sector job depend on which sample is being considered. In the sample containing self-employed, being in the US Border increases the probability of applying for such jobs, while being in the other regions decreases it. In the sample with wage workers only however, all the regions have large positive effects for applying for such jobs, in comparison to Mexico City.

The effects of wealth proxies on the decision of applying for a formal sector job are similar to the ones found in the univariate (free entry) model, namely, they positively affect the desire of being in the formal sector.[18]

The decision of the employer to hire a worker is always positively associated with the education of the applicant. Younger workers are more likely to be accepted into formal sector jobs in the sample including self-employed, but in the sample including wage workers only, age plays either no role or a negative one. This could be an indication that old workers who are rejected from formal sector jobs tend to become informally self-employed. Being a male is negatively associated with the probability of being hired in the sample including self-employed, but shows no effect in the sample with wage workers.

Finally, regarding the impact of region dummies on the probability of being accepted in the formal sector, while no clear pattern arises in the sample with self-employed, the other sample consistently shows that applicants to formal sector jobs are more likely to be hired in the US Border region, while workers in other regions have a smaller relative probability of being accepted in that sector.

[18]With the exception of the "Type of Dwelling" variable.

Table 5.3: Probit Models of Sector Allocation. Wage Workers Only. Q1:87-Q2:93

	No Rationing Model		Rationing Model			
	Pr(Formal=1)		*Pr(A=1)*		*Pr(E=1|A=1)*	
Age						
Linear	0.0486	***	0.0554	***	0.0074	
	(0.007)		(0.005)		(0.013)	
Squared	-0.0005	***	-0.0005	***	-0.0002	
	(0.000)		(0.000)		(0.000)	
Education						
Linear	0.1383	***	0.1034	***	0.0997	***
	(0.007)		(0.004)		(0.015)	
Squared	-0.0011	***	-0.0003		0.0079	
	(0.000)		(0.000)		(0.002)	
Male	0.0660	***	-0.4319	***	5.5103	
	(0.018)		(0.030)		(143.8)	
Region						
Mexico City (omitted)						
US Border	0.1575	***	0.1048	***	0.3659	***
	(0.020)		(0.023)		(0.071)	
North	0.1855	***	0.2984	***	-0.3903	***
	(0.017)		(0.016)		(0.038)	
Center	0.0534	***	0.0849	***	-0.1843	***
	(0.017)		(0.014)		(0.036)	
South	0.1760	***	0.1841	***	0.0001	
	(0.026)		(0.040)		(0.110)	
Married	0.1739	***	0.2141	***		
	(0.020)		(0.014)			
Household Structure						
# hh members >11yrs	0.0058		0.0015			
	(0.004)		(0.003)			
# hh members <12yrs	-0.0298	***	-0.0351	***		
	(0.006)		(0.004)			
Wealth Proxies						
Cluster Average Income	0.0000	***	0.0000	***		
	(0.000)		(0.000)			
Constant	-1.2416	***	-0.7500	***	0.2080	
	(0.150)		(0.102)		(0.268)	
No. Observations	143,951				143,951	
Log-Likelihood	-47,377.9				-46,708.9	
LRT statistic	1338.0		p-value		0.000	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.4: Probit Models of Sector Allocation. Wage Workers Only. Q3:94-Q4:01.

	Q3:94-Q1:99			*Q2:99-Q4:01*		
	No Rationing Model	Rationing Model		No Rationing Model	Rationing Model	
	Pr(Formal=1)	*Pr(A=1)*	*Pr(E=1|A=1)*	*Pr(Formal=1)*	*Pr(A=1)*	*Pr(E=1|A=1)*
Age						
Linear	0.0186 **	0.0249 ***	-0.0214	0.0178 *	0.0133	0.0320 **
	(0.008)	(0.006)	(0.021)	(0.010)	(0.007)	(0.014)
Squared	-0.0001	-0.0002 ***	0.0002	-0.0001	-0.0001	-0.0004 **
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Linear	0.0898 ***	0.0591 ***	0.0580 ***	0.0875 ***	0.0662 ***	0.1217 ***
	(0.007)	(0.005)	(0.022)	(0.009)	(0.007)	(0.012)
Squared	0.0007 *	0.0014 ***	0.0108 ***	0.0011 **	0.0015 ***	0.0007
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.001)
Male	0.0529 ***	-0.2428 ***	5.4758	0.0362	-0.5981 ***	5.4950
	(0.020)	(0.022)	(134.7)	(0.023)	(0.085)	(122.8)
Region						
Mexico City(omitted)						
US Border	0.5332 ***	0.5119 ***	0.4350 ***	0.6346 ***	0.6389 ***	0.4076 ***
	(0.028)	(0.028)	(0.115)	(0.036)	(0.036)	(0.069)
North	0.2380 ***	0.3697 ***	-0.5730 ***	0.4020 ***	0.5561 ***	-0.1161 *
	(0.021)	(0.019)	(0.056)	(0.027)	(0.025)	(0.063)
Center	0.0452 **	0.0934 ***	-0.3521 ***	0.1909 ***	0.3322 ***	-0.2499 ***
	(0.021)	(0.017)	(0.056)	(0.026)	(0.023)	(0.055)
South	0.2477 ***	0.3345 ***	-0.4358 ***	0.3125 ***	0.3966 ***	-0.0092
	(0.033)	(0.047)	(0.118)	(0.041)	(0.056)	(0.101)
Married	0.1503 ***	0.1850 ***		0.1536	0.2275 ***	
	(0.022)	(0.015)		(0.025)	(0.022)	
Household Structure						
# hh members >11yrs	-0.0100 **	-0.0119 ***		-0.0068	-0.0151 ***	
	(0.005)	(0.003)		(0.006)	(0.005)	
# hh members <12yrs	-0.0154 **	-0.0231 ***		-0.0038	-0.0074	
	(0.008)	(0.005)		(0.010)	(0.007)	
Wealth Proxies						
Cluster Average Income	0.0000 ***	0.0000 ***		0.0000	0.0000 ***	
	(0.000)	(0.000)		(0.000)	(0.000)	
Type of Dwelling	-0.0065	-0.0096		0.0195	0.0203 *	
	(0.015)	(0.009)		(0.018)	(0.011)	
Solid Walls	0.0505	0.0653 **		0.0101	0.0275	
	(0.034)	(0.029)		(0.046)	(0.041)	
Solid Roofs	0.0900 ***	0.1007 ***		0.1228 ***	0.1210 ***	
	(0.029)	(0.017)		(0.036)	(0.023)	

Table 5.4 (Continued)

Solid Floors	0.1475	***	0.1475	***			0.1266	***	0.1428	***		
	(0.021)		(0.014)				(0.026)		(0.018)			
Electricity	0.1181		0.1264				0.1103		0.1099			
	(0.108)		(0.092)				(0.171)		(0.149)			
Water	0.0229		0.0289				0.0954		0.1277	**		
	(0.052)		(0.037)				(0.078)		(0.055)			
Sewage	0.1153	***	0.1220	***			0.0295		0.0502			
	(0.044)		(0.031)				(0.056)		(0.050)			
Phone	0.1366	***	0.1318	***			0.0969	***	0.1022	***		
	(0.022)		(0.014)				(0.025)		(0.017)			
Other services	0.0109		0.0164				0.1074	***	0.1454	***		
	(0.029)		(0.018)				(0.031)		(0.023)			
Kitchen	0.1411	***	0.1567	***			0.1026	***	0.1228	***		
	(0.027)		(0.015)				(0.031)		(0.019)			
Constant	-1.2227	***	-0.9642	***	0.9669	**	-1.2976	***	-0.6448	***	-0.5296	*
	(0.199)		(0.148)		(0.429)		(0.265)		(0.220)		(0.282)	
No. Observations	100,637				100,637		71,767				71,767	
Log-Likelihood	-38,224.0				-37,864.9		-27,111.6				-26,859.3	
LRT statistic	718.2		p-value		0.000		504.7		p-value		0.000	

* p < 0.1, ** p < 0.05, *** p < 0.01

5.4.2 Earnings Equations

The results for the estimated earnings equations for each sector, corrected for selectivity bias, are included in Tables 5.5 and 5.6. These tables show that earnings rise with age at a decreasing rate, and the age profile is less steep in the wage worker's sample. For most of the panels, education displays a U-pattern on earnings. In the sample including (informal) self-employed, an extra year of education brings higher gains to informal sector workers, holding everything else constant; however, the opposite occurs in the sample including only wage workers. This result might reflect that there are some returns to capital in what self-employed individuals report as their "labor earnings". The effect of being male on earnings is quite high, and in the wage worker's sample, this effect is higher in the informal sector.

Being in the US Border region brings higher earnings than any other region, while being in the South lowers earnings more than any other region. Furthermore, *ceteris paribus*, being in the informal sector in the US Border region brings higher earnings than being in the formal sector.

Finally, the selectivity correction parameters for the workers and the employers equation, i.e., λ_w, λ_e, are for the most part negative and statistically significant. This is somewhat puzzling, because it means that the unobserved factors that lead to higher earnings (in each sector) are negatively correlated with the unobserved factors that lead workers to apply to formal sector jobs *and* the unobserved factors that lead formal sector employers to accept applicants.

Table 5.5: Earnings Equations. Informal Self-employed Included.
Dep.Var. Log-earnings.

	Q1:87-Q2:93		*Q3:94-Q1:99*		*Q2:99-Q4:01*	
	Formal Workers	Informal Sector	Formal Workers	Informal Sector	Formal Workers	Informal Sector
Age						
Linear	0.0505 ***	0.0669 ***	0.0579 ***	0.0723 ***	0.0508 ***	0.0556 ***
	(0.003)	(0.005)	(0.005)	(0.006)	(0.006)	(0.006)
Squared	-0.0005 ***	-0.0007 ***	-0.0005 ***	-0.0008 ***	-0.0004 ***	-0.0006 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Linear	-0.0166	0.0073	-0.0831 ***	-0.0281 ***	-0.1188 ***	-0.0205 ***
	(0.011)	(0.005)	(0.013)	(0.006)	(0.017)	(0.006)
Squared	0.0034 ***	0.0023 ***	0.0054 ***	0.0035 ***	0.0064 ***	0.0037 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.3211 ***	0.5903 ***	0.4945 ***	0.4698 ***	0.5696 ***	0.5570 ***
	(0.036)	(0.026)	(0.046)	(0.026)	(0.049)	(0.027)
Region						
Mexico City (omitted)						
US Border	0.3128 ***	0.4116 ***	0.0611 **	0.3182 ***	0.0201	0.4410 ***
	(0.015)	(0.015)	(0.025)	(0.021)	(0.035)	(0.022)
North	0.0178 *	0.0413 ***	-0.1058 ***	-0.0488 ***	-0.1441 ***	0.0930 ***
	(0.011)	(0.014)	(0.025)	(0.018)	(0.034)	(0.019)
Center	0.0644 ***	0.0717 ***	0.0171	-0.0271	-0.0859 ***	0.0000
	(0.012)	(0.015)	(0.017)	(0.017)	(0.022)	(0.017)
South	-0.0976 ***	-0.0491 **	-0.2732 ***	-0.1897 ***	-0.4237 ***	-0.2863 ***
	(0.014)	(0.021)	(0.037)	(0.025)	(0.047)	(0.031)
Selectivity Correction						
λ_w	-0.2720 ***	-0.1265 *	-0.5531 ***	0.0597	-0.5176 ***	-0.1025
	(0.027)	(0.077)	(0.038)	(0.089)	(0.041)	(0.107)
λ_e	-0.6254 ***	-0.3372 ***	-1.7858 ***	-0.5222 ***	-2.1249 ***	-0.3987 ***
	(0.187)	(0.040)	(0.248)	(0.057)	(0.275)	(0.058)
Constant	6.9253 ***	5.5630 ***	7.5435 ***	5.3851 ***	8.1473 ***	5.7394 ***
	(0.125)	(0.110)	(0.173)	(0.125)	(0.223)	(0.136)
R-squared	0.305	0.276	0.373	0.280	0.384	0.315
No. Obs.	118,580	65,990	78,922	48,878	56,494	32,569

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.6: Earnings Equations. Wage Workers Only.
Dep.Var. Log-earnings.

	Q1:87-Q2:93		*Q3:94-Q1:99*		*Q2:99-Q4:01*	
	Formal Workers	Informal Sector	Formal Workers	Informal Sector	Formal Workers	Informal Sector
Age						
Linear	0.0306 ***	0.0303 ***	0.0277 ***	0.0341 ***	0.0279 ***	0.0249 ***
	(0.003)	(0.009)	(0.005)	(0.008)	(0.004)	(0.008)
Squared	-0.0003 ***	-0.0004 ***	-0.0003 ***	-0.0005 ***	-0.0003 ***	-0.0003 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Education						
Linear	-0.0454 ***	0.0137	-0.0659 ***	-0.0376 ***	-0.0761 ***	-0.0258 ***
	(0.006)	(0.010)	(0.006)	(0.008)	(0.006)	(0.009)
Squared	0.0042 ***	-0.0003	0.0050 ***	0.0015 ***	0.0056 ***	0.0006 ***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
Male	0.2921 ***	0.4451 ***	0.2953 ***	0.4796 ***	0.3169 ***	0.5865 ***
	(0.010)	(0.053)	(0.012)	(0.043)	(0.015)	(0.055)
Region						
DF(omitted)						
US Border	0.2353 ***	0.3250 ***	0.0024	0.1492 ***	0.0169	0.1848 ***
	(0.011)	(0.021)	(0.019)	(0.030)	(0.019)	(0.035)
North	-0.0407 ***	-0.0032	-0.1420 ***	-0.1417 ***	-0.1201 ***	-0.0977 ***
	(0.011)	(0.020)	(0.019)	(0.026)	(0.018)	(0.028)
Center	0.0147	0.0185	-0.0826 ***	-0.0979 ***	-0.1310 ***	-0.1037
	(0.009)	(0.019)	(0.015)	(0.022)	(0.016)	(0.023)
South	-0.1622 ***	-0.1706 ***	-0.3015 ***	-0.3537 ***	-0.3542 ***	-0.4366 ***
	(0.016)	(0.027)	(0.028)	(0.034)	(0.025)	(0.041)
Selectivity Correction						
λ_w	-1.2166 ***	-0.2503 ***	-1.4942 ***	-0.4445 ***	-1.3083 ***	-0.5075
	(0.091)	(0.051)	(0.077)	(0.049)	(0.068)	(0.052)
λ_e	-0.4322 ***	-0.0852	-0.5117 ***	-0.3233 ***	-0.6989 ***	-0.5353 ***
	(0.034)	(0.087)	(0.046)	(0.085)	(0.066)	(0.070)
Constant	7.6048 ***	6.4468 ***	7.8484 ***	6.2513 ***	7.8816 ***	6.3159 ***
	(0.095)	(0.172)	(0.118)	(0.144)	(0.109)	(0.160)
R-squared	0.308	0.229	0.388	0.222	0.397	0.261
No. Obs.	118,580	17,401	78,922	13,967	56,494	9,354

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.4.3 Predictions of the Model

This section analyzes some of the predictions stemming from the previous models.

In Table 5.7 the number of predicted restricted individuals as a percentage of the total population is presented. The specification including informal self-employed is labeled Model 1 from now on, and the one keeping only wage workers, Model 2. The bottom table also contains the same numbers expressed as a percentage of the total number of individuals in each sector. In order to make these predictions it is assumed that any individual currently working in the informal sector, for whom the probability of applying to a formal sector job is greater than 0.5, i.e., $\Pr(A=1|X)>0.5$, is restricted from entering the formal sector.

These two tables show that Models 1 and 2 predict most of the informal sector individuals to be rationed out of the formal sector. In both cases more than 90% of the individuals in the informal sector are predicted to be rationed. This number seems extremely high. Although, it makes sense to believe that the segmentation model fits better the Mexican case based on the results presented in the previous sections, there is also evidence that a fraction of the informal self-employed are in that state by choice and not by force.[19]

The fact that the model predicts such a high number of restricted individuals, calls for further revisions of the econometric specification. As a defense of the model and the segmentation result obtained in the previous sections, it must be said that these extreme predictions do not disqualify the previous findings in the probit model. After all, a probit model is not designed to maximize the prediction

[19] Maloney (1999) provides evidence based on a survey on micro-enterprises that 2/3rds of the workers leaving a formal sector wage job to form an enterprise do so looking for either higher wages or more independence. The problem with his result is that it is based on an extremely small sample (less than 200 obs.), and it does not distinguish between formal and informal self-employed individuals.

Table 5.7: Percentage of Predicted Restricted Individuals.

% of Total Population		
	Model 1	Model 2
Period	%	%
Q1:87-Q2:93	31.4	12.5
Q3:94-Q1:99	35.7	15.9
Q2:99-Q4:01	35.3	15.6

	% of Individuals in each Sector		
	Model 1		Model 2
Period	% Self-employed	% Workers	%Workers
Q1:87-Q2:93	91.5	89.6	99.9
Q3:94-Q1:99	93.4	91.2	98.1
Q2:99-Q4:01	95.0	93.8	96.4

Model 2 excludes Informal Self-employed.

quality of the model (as for instance Manski's Maximum Score model). Furthermore, the model does not incorporate panel information on its structure to control for potential fixed unobserved heterogeneity. Two ways to test the robustness of the sector choice model are: (1) perform tests of normality, and (2) allow for heteroskedasticity in the probit models.[20] These extensions will be left for future research. Instead, the approach taken in the rest of the chapter is to provide some further evidence on the sector and earnings mobility of these (predicted) restricted individuals using the panel information available in the survey.

Table 5.8 contains the percentage of individuals, restricted and non-restricted, who moved into the formal sector. This table also displays the percentage of restricted individuals who moved into the formal sector during a year, by their sector of origin.[21] The interesting fact in this table is that around 30% of the predicted

[20] The presence of heteroskedasticity in nonlinear models alters not only the estimates of the standard errors, but also of the main parameters of interest.

[21] The latter calculation is made for Model 1 only, since this is the only model that contains two types of informal sector workers (the self-employed and the wage workers).

Table 5.8: Fraction of Individuals who Moved into Formal Sector.

	Model 1		Model 2	
Period	Restricted	Non Restricted	Restricted	Non Restricted
Q1:87-Q2:93	23.2	76.4	35.5	79.6
Q3:94-Q1:99	22.8	83.4	33.5	86.5
Q2:99-Q4:01	24.9	83.9	35.5	85.8

Period	% Restricted Wage Workers	% Restricted Self-employed
Q1:87-Q2:93	37.7	17.8
Q3:94-Q1:99	34.8	17.6
Q2:99-Q4:01	36.0	19.7

Estimates based on Model 1

restricted individuals manage to enter the formal sector within one year.[22] This result, rather than undermine the conclusions of the model, it indicates that maybe the high rates of predicted restricted individuals arise because these predictions are made for a given point in time (the first period of the interview). However, as time goes by restricted individuals manage to move into the formal sector. This finding by the way, does not contradict the theory of segmented labor markets. On the contrary, the idea that workers might move one day into the formal sector lies at the core of models like the ones of Harris and Todaro (1970) and Fields (1975). It is precisely the positive probability of one day becoming formal that motivates workers to stay in the urban areas, even in the presence of unemployment.[23]

Moving now to the analysis of earnings mobility, Table 5.9 shows the potential

[22]The fact that the vast majority of the non-restricted individuals is found in the formal sector later on is not surprising, given that most of these workers started in the formal sector and kept their jobs over time.

[23]A subsidiary result obtained from this exercise is that restricted individuals have a much higher rate of attrition from the panel than non-restricted ones, even after controlling for a set of individual characteristics. This is consistent with the idea that restricted individuals would move to other locations in search of better opportunities.

Table 5.9: Potential Median Earnings Gains for a Restricted Individual Moving into Formal Sector.

	Mx Pesos (2002=100)		% of Median Earnings	
Period	Model 1	Model 2	Model 1	Model 2
Q1:87-Q2:93	221.0	519.8	7.1	22.8
Q3:94-Q1:99	472.4	537.1	20.2	29.0
Q2:99-Q4:01	436.1	501.6	16.7	23.8

Model 2 excludes Informal Self-employed.

gains that a restricted individual in the informal sector could experience by moving to the formal sector. These potential gains are constructed by predicting formal sector earnings for restricted individuals, using the selectivity corrected earnings equations (5.8) and (5.9), and subtracting from this prediction their current earnings. Both models predict that there will be positive median gains from such transition (fluctuating between 7% and 30% of the initial period median earnings). In particular, the model including only wage workers (Model 2) predicts even higher gains from this transition, because wage workers earn less than the informal self-employed.

Finally, Table 5.10 presents the actual earnings mobility experienced by the restricted individuals one year after the period in which the model was fitted. The table decomposes the results by sector of destination, and includes both unconditional gains and gains conditional on a set of individual characteristics.[24] This table show mobility patterns similar to the ones presented in Chapter 4. Both conditionally and unconditionally, restricted individuals experience higher mobility when moving into formal self-employment, and besides movements into unemployment, which trivially imply large losses, the worst destination sector in terms of earnings mobility is the informal wage work. In most of the cases, transitions

[24]The control variables were age, education, gender, regional dummies and initial earnings.

into informal self-employment bring larger gains than the ones into formal wage work.

It is important to remark that without panel data that follows individuals for longer periods of time, it is hard to assess whether these sector transitions are a permanent phenomenon or a transitory one; i.e., it is hard to know if an informal worker who moved to the formal sector will keep his job for a long time. Also, it is important to note that the actual unconditional gains of individuals moving into the formal sector are smaller than the gains predicted by the earnings equations (5.8) and (5.9). This is an indication that the model overestimates the potential gains of moving into the formal sector.

Table 5.10: Actual Yearly Earnings Mobility of Restricted Individuals by Sector of Destination

			Unconditional Medians			
		Model 1			Model 2	
	Q1:87-Q2:93	Q3:94-Q1:99	Q2:99-Q4:01	Q1:87-Q2:93	Q3:94-Q1:99	Q2:99-Q4:01
Unemployed	-2997.6	-2325.9	-2268.6	-2413.1	-2136.5	-1803.3
Informal Worker	-77.1	-236.6	-55.3	2.9	-162.5	82.1
Informal Self-emp	39.4	-140.8	94.0	506.8	183.4	460.1
Formal Self-emp	1498.8	1811.3	2510.8	5052.9	2176.1	4063.9
Formal Worker	83.8	23.0	304.6	209.2	56.6	354.7

			Conditional on X*			
		Model 1			Model 2	
	Q1:87-Q2:93	Q3:94-Q1:99	Q2:99-Q4:01	Q1:87-Q2:93	Q3:94-Q1:99	Q2:99-Q4:01
Unemployed	-8683.9	-9065.6	-5712.0	-4820.1	-3019.1	-3915.6
Informal Worker	-5034.7	-6214.4	-2737.2	-1923.8	-755.8a	-1374.4
Informal Selfemp	-3745.8	-5255.1	-1939.0	-836.6a	-107.2a	-688.8a
Formal Selfemp	2353.5a	-1808.1a	3864.4	2479.1	4659.1	4071.6
Formal Worker	-4705.1	-6003.0	-2213.4	-1512.0	-524.6a	-1000.5a

* Model Estimated via Least Squares.
a. Statistically insignificant different from 0 at 90%.
Model 2 excludes Informal Self-employed.
All the numbers are in real Mx Pesos of 2002.

5.5 Conclusions

This chapter tested whether Mexican labor markets are segmented, in particular whether formal sector jobs are rationed, and what are the implications of this segmentation for earnings mobility.

The test for segmentation in labor markets consisted of comparing a model that assumed free entry into the formal sector to one that allowed for the possibility of rationing of formal sector jobs. Since the data only reveals the sector of the worker, and not his willingness to enter the formal sector, a bivariate probit with partial observability was used when estimating the components of the sector allocation model that allowed for rationing.

In addition to this, selectivity-corrected earnings equations were estimated for each sector. These equations were used to calculate the potential earnings gains of entering the formal sector, for individuals predicted to be restricted from entering this sector. Finally, the chapter used the panel structure of the data to analyze the sector and earnings mobility experienced in further periods, by individuals predicted to be restricted from entering the formal sector in the initial period.

To the question of whether formal salaried jobs are rationed in Mexico, the answer obtained was affirmative. The discrete choice model that assumed free movement across sectors was strongly rejected versus a richer model that allowed the possibility of formal sector job rationing. This conclusion was reached for samples with and without informal self-employed.

Some important factors affecting the worker's decision to apply to formal sector jobs were the individual's years of education and his wealth proxies, with a positive effect, and the number of children in the household, with a negative effect. Other variables like age, gender, marital status, region and number of adults present in the household played a role too in affecting this decision, but their impact varied depending on the years and the sample analyzed (meaning by this whether the estimations included informal self-employed or not).

Regarding the employer's decision, more years of education for the worker always increased his probability of being hired in the formal sector. Also, being a woman and being young increased this probability in the sample containing informal self-employed individuals. Finally, in the sample with wage workers only, living in cities along the US Border increased the probability of being hired by a formal sector employer.

The selectivity-corrected earnings equations showed positive effects from age, and large positive effects on earnings for being male. Education presented a U-shaped pattern, first decreasing then increasing, and in the sample with informal wage workers only, more education brought larger gains in the formal sector. Also, holding everything else constant, living in cities along the US Border was associated with higher earnings, especially in the informal sector.

The sector choice model predicted that more than 90% of the informal sector workers were restricted from entering the formal sector. Although this number seems too high, analysis of the sectoral mobility patterns in further periods showed that around 30% of those restricted individuals managed to find a formal sector job within one year. This suggests that part of the reason why the model predicted such a high number of restricted individuals is because the predictions hold at a given point in time and, as time passes by, sector mobility starts to occur.

To the question of "What are the potential earnings gains that rationed individuals could experience by moving into the formal sector?" the estimations predict that if individuals rationed out of the formal sector were to move to the formal sector, they would experience gains going from 7% to 30% of their current earnings, depending on the sample. While the prediction of upward earnings mobility is confirmed by looking at the actual earnings gains experienced by restricted individuals who moved into the formal sector after a year, the actual earnings gains were smaller than the predicted ones. The most desirable destination sector in terms of earnings mobility is formal self-employment, and the least desirable one is informal salaried work.

Although the proposed model performs better than a free entry one, many issues require further research. In particular, tests for the distributional assumptions, as well as for heteroskedasticity in the data, should be performed. It is also important to explore whether under less restrictive assumptions, the model generates more credible predictions about the extent of segmentation in labor markets. Even with these limitations, the previous estimations show that the issue of segmentation in Mexican labor markets is far from settled, as some previous studies had maintained.

5.6 Appendix

5.6.1 Descriptive Statistics

Since the models in this chapter are cross-sectional, they are estimated at the first interview for each individual. This eliminates the potential negative effects of attrition from the panel, and makes the results representative of the urban population in Mexico. Since using individuals observed at the first interview imposes less restrictions than the ones previously encountered in chapters 3 and 4, the sample analyzed here differs somewhat from the previous ones. Due to this discrepancy, new descriptive statistics for this less restrictive sample are presented in Table 5.11.

For the purpose of comparison with chapters 3 and 4, the sample was restricted to individuals 25 to 60 years old.

Table 5.11: Descriptive Statistics for the Sample of Individuals Analyzed in Chapter 5.

	Q1:87-Q2:93				*Q3:94-Q1:99*				*Q2:99-Q4:01*			
	Mean	s.e.(mean)	Median	No. Obs	Mean	s.e.(mean)	Median	No. Obs	Mean	s.e.(mean)	Median	No. Obs
Earnings	4692.3	13.034	3288.3	185274	4310.2	19.652	2873.0	128523	4578.7	18.104	3158.2	89449
Age	37.65	0.021	36	197221	38.16	0.025	37	139685	38.40	0.030	37	97570
Education	8.99	0.011	9	197221	9.77	0.013	9	139685	10.08	0.015	9	97570
Male	0.68	0.001	1	197221	0.66	0.001	1	139685	0.64	0.002	1	97570
Region												
DF	0.6	0.0	1.0	197221	0.5	0.0	1.0	139685	0.5	0.0	1.0	97570
US Border	0.06	0.001	0	197221	0.07	0.001	0	139685	0.08	0.001	0	97570
North	0.17	0.001	0	197221	0.18	0.001	0	139685	0.18	0.001	0	97570
Center	0.20	0.001	0	197221	0.20	0.001	0	139685	0.20	0.001	0	97570
South	0.02	0.000	0	197221	0.02	0.000	0	139685	0.02	0.000	0	97570
Married	0.73	0.001	1	197221	0.71	0.001	1	139685	0.71	0.001	1	97570
Household Structure												
# hh members >11yrs	3.85	0.005	3	197221	3.69	0.005	3	139685	3.54	0.005	3	97570
# hh members <12yrs	1.30	0.003	1	197221	1.12	0.003	1	139685	1.05	0.004	1	97570
Sector												
Informal Salaried Worker	0.09	0.001	0	197221	0.12	0.001	0	139685	0.12	0.001	0	97570
Informal Self-employed	0.25	0.001	0	197221	0.27	0.001	0	139685	0.25	0.001	0	97570
Formal Self-employed	0.00	0.000	0	197221	0.00	0.000	0	139685	0.00	0.000	0	97570
Formal Salaried worker	0.66	0.001	1	197221	0.62	0.001	1	139685	0.63	0.002	1	97570
Wealth Proxies												
Cluster Average Income	4090.8	4.687	3495.5	197221	3474.8	5.172	2881.3	139685	4070.4	6.789	3467.4	97570
Type of Dwelling	.	.	.	.	1.43	0.002	1	139685	1.45	0.002	1	97570
Solid Walls	.	.	.	.	0.96	0.001	1	139685	0.96	0.001	1	97570
Solid Roofs	.	.	.	.	0.85	0.001	1	139685	0.87	0.001	1	97570
Solid Floors	.	.	.	.	0.57	0.001	1	139685	0.59	0.002	1	97570
Electricity	.	.	.	.	1.00	0.000	1	139685	1.00	0.000	1	97570
Water	.	.	.	.	0.98	0.000	1	139685	0.99	0.000	1	97570
Sewage	.	.	.	.	0.97	0.000	1	139685	0.98	0.000	1	97570
Phone	.	.	.	.	0.54	0.001	1	139685	0.60	0.002	1	97570
Other services	.	.	.	.	0.20	0.001	0	139685	0.24	0.001	0	97570
Kitchen	.	.	.	.	0.84	0.001	1	139685	0.83	0.001	1	97570

5.6.2 Selectivity Correction Model

This section presents the derivation of the selectivity correction terms in its most general form, in a model of double selection with partial observability. The notation in this section is different from the one previously used, in order to keep the model general.

Let the two selection rules be determined by the following equations

$$y_{i1}^* = Z_{i1}\gamma_1 + u_{i1} \tag{5.10}$$

$$y_{i2}^* = Z_{i2}\gamma_2 + u_{i2} \tag{5.11}$$

The terms y_{i1}^*, y_{i2}^* are the latent variables that determine each selection rule, the vectors Z_{i1}, Z_{i2} are the observable factors affecting these latent variables, and the error terms u_{i1}, u_{i2} are assumed to be bivariate normally distributed with correlation ρ among themselves, i.e., $(u_{i1}, u_{i2})' \sim N(0, \Sigma)$.

The outcome of this latent variables is summarized by two dichotomous variables D_{i1} and D_{i2}, where

$$D_{i1} = \begin{cases} 1 & \text{if } y_{i1}^* \geq 0, \\ 0 & \text{if } y_{i1}^* < 0. \end{cases} \qquad D_{i2} = \begin{cases} 1 & \text{if } y_{i2}^* \geq 0, \\ 0 & \text{if } y_{i2}^* < 0. \end{cases}$$

In this model of partial observability the individual is observed in state A if and only if both $y_{i1}^* > 0$ and $y_{i2}^* > 0$, i.e., if $D_{i1} = 1$ and $D_{i2} = 1$. Denote the probability of this event by P_A, i.e.,

$$P_A = Pr(D_1 = 1, D_2 = 1) = G(Z_1\gamma_1, Z_2\gamma_2; \rho)$$

where $G(\cdot, \cdot; \rho)$ is the bivariate normal distribution function with correlation coefficient ρ. The probability of the individual being in the other state B, is $1 - P_A$.

There are two outcome equations associated with each state (A,B)

$$y_{iA} = X_{iA}\beta_A + \varepsilon_{iA} \tag{5.12}$$

$$y_{iB} = X_{iB}\beta_B + \varepsilon_{iB} \tag{5.13}$$

The error terms ε_{iA} and ε_{iB} are assumed to be equal to each other, i.e. $\varepsilon_{iA} = \varepsilon_{iB} = \varepsilon_i$.[25] Furthermore, the error term ε_i is assumed to have mean zero, finite variance σ^2_ε, and to be a linear function of u_{i1}, u_{i2}.[26]

For simplicity of notation denote $C_1 = Z_1\gamma_1, C_2 = Z_2\gamma_2$, and define

$$C_1^* = \frac{C_1 - \rho C_2}{(1-\rho^2)^{1/2}} \qquad\qquad C_2^* = \frac{C_2 - \rho C_1}{(1-\rho^2)^{1/2}}.$$

Poirier (1980) and Tunalı (1986) show that for equation (5.12) the selectivity correction terms are

$$\lambda_{A1} = \frac{\phi(C_1)\Phi(C_2^*)}{P_A} \qquad\qquad \lambda_{A2} = \frac{\phi(C_2)\Phi(C_1^*)}{P_A}$$

i.e., the conditional expectation of ε_i given that the individuals are observed in state A is

$$E(\varepsilon_i|D_1 = 1, D_2 = 1) = \rho_{A1}\lambda_{A1} + \rho_{A2}\lambda_{A2}$$

.

Using these results, equivalent terms can be derived for equation (5.13) by applying the Law of Iterated Expectations (LIE). In particular, it follows that

$$\lambda_{B1} = \frac{-\phi(C_1)\Phi(C_2^*)}{1-P_A} \qquad\qquad \lambda_{B2} = \frac{-\phi(C_2)\Phi(C_1^*)}{1-P_A}$$

[25]This assumption will be necessary to be able to apply the Law of Iterated Expectations.

[26]This linearity assumption together with the bivariate normality of (u_{i1}, u_{i2}) is equivalent to assuming trivariate normality of the vector $(u_{i1}, u_{i2}, \varepsilon_i)$. However, if the bivariate normality assumption is dropped, but the linearity one is kept, this can lead to other types of two-step selectivity correction models.

since the LIE implies

$$E(\varepsilon_i) = P_A E(\varepsilon_i | D_1 = 1, D_2 = 1) + (1 - P_A) E(\varepsilon_i | D_1 = 0 \vee D_2 = 0)$$

and since it was assumed that $E(\varepsilon_i) = 0$. With this expression in hand it follows that

$$E(\varepsilon_i | D_1 = 0 \vee D_2 = 0) = \rho_{B1}\lambda_{B1} + \rho_{B2}\lambda_{B2}.$$

Extensions for the expressions for the conditional second moments and variance of ε_i, as well as for the standard error of the regression, the variance-covariance matrix, and the gradient vectors that apply to equation (5.13) readily follow from the expressions included in Tunalı (1986), by substituting the ρ_{A1}, ρ_{A2} by ρ_{B1}, ρ_{B2}, and P_A by $1 - P_A$, where appropriate. Note also that the expressions in (5.6) and (5.7) follow from the previously derived λ's by assuming $\rho = 0$.

Chapter 6

Conclusions

This dissertation studied labor earnings mobility in the short-run and the structure of labor markets in urban Mexico, from 1987 to 2002. The topics addressed included aggregate earnings mobility, the determinants of earnings mobility with a special emphasis on the impact of initial earnings on mobility, and segmentation in Mexican labor markets and its implications for earnings mobility. These topics were analyzed under varying macroeconomic conditions. In particular, the effects of the 1994 Peso crisis were examined.

The first substantive chapter on this dissertation (Chapter 3) was devoted to issues of aggregate earnings mobility. In particular, it focused on measuring the concepts of Directional mobility and Mobility as an equalizer of longer-term earnings. Regarding Directional mobility, the questions asked were: "What are the average earnings gains and losses in the economy?" and "Are these earnings mobility patterns the same for different groups of the population?". The groups of the population analyzed included age, education, gender, initial earnings quintile, region and sector groups. In overall terms, the results obtained show that average earnings mobility fluctuated around zero, with the exception of the late eighties and early 2000, when individuals experienced upward earnings mobility, and of the years following the 1994 Peso crisis, when individuals experienced large losses. These patterns were shared by the majority of the subgroups of the population. The only groups for which significant differences appear in their mobility patterns are initial earnings quintile and sector groups. In both cases, the most advantaged individuals experienced the largest losses, while the ones that initially had the smallest earnings, experienced the largest gains. These results held both in absolute and proportional terms.

The questions concerning Mobility as an equalizer of longer-term earnings were: "Does mobility equalize earnings over time?", "Does mobility equalize earnings *within* groups over time?" and "Does mobility equalize earnings *between* groups over time?". The data showed that mobility equalized longer-term earnings for most of the periods studied. Only for a couple of years in the late eighties, a disequalizing pattern was found. It was also found that in general, mobility equalized longer-term earnings *within-groups* over a year, while it sometimes equalized and others disequalized longer-term earnings *between-groups* of the population. The only exception to this occurred for initial earnings quintile groups, for which earnings mobility equalized longer-term earnings between-groups. One interesting finding was that mobility almost always disequalized longer-term earnings between men and women, but it reduced the longer-term earnings inequality within-genders. At the end, the equalizing effect within genders dominated the disequalizing one found between-genders.

The second substantive chapter in this dissertation is Chapter 4, which examined earnings mobility at the individual level. In particular, it studied the impact of initial earnings and other variables on earnings mobility. In this chapter, special attention was given to the issue of whether mobility benefited initially advantaged individuals, or whether it benefited the disadvantaged ones. The answer to the question "Are the most advantaged individuals gaining more (losing less) in terms of earnings changes?" is that no; in most of the cases the advantaged individuals kept their advantage, but the rich individuals did not get richer. Although the comparison of earnings changes to initial earnings showed a high level of convergence between the earnings of rich and poor, further analysis revealed that part of this convergence was due to the adjustment of earnings from transitory fluctuations to its permanent level. In other words, besides the effects of transitory adjustments in earnings, there was little, or no convergence, between rich and poor. The only major exception to this finding occurred in the aftermath of the 1994 Peso crisis, when individuals with a high permanent advantage experienced greater losses than everybody else. This crisis generated proportional losses across the earnings distribution, leading to larger losses in pesos for the permanent rich individuals. In the light of these results, the conclusions reached in Chapter 3 on the role of mobility in equalizing longer-term earnings between initial earnings quintile groups, must be reinterpreted as arising most of the times because of transitory adjustments in earnings, and not because of a lasting convergence between rich and poor.

Chapter 4 also addressed the *ceteris paribus* impact of socioeconomic characteristics of the individual on earnings mobility, and how accounting for these socioeconomic factors affects the impact of initial earnings on mobility. Here it was found that, holding everything else constant, more education lead to negative mobility for individuals with a low level of education, however after a certain point (around elementary education), more education was associated with upward conditional mobility. Being male and living in cities by the US Border and in the North of the country, brought large positive earnings mobility. Sector transitions into formal self-employment were associated with the largest gains (smallest losses), while transitions into informal wage work brought about the largest losses (smallest gains) (besides, of course, movements into unemployment). Finally, once all these socioeconomic factors were controlled for, initial earnings had a strong negative effect on mobility, meaning that individuals converged to their own permanent earnings. The fact that the conditional convergence rates (the convergence rates obtained after controlling for socioeconomic characteristics of the individual) were stronger than the unconditional ones (without such controls), implies that the overall effect of these characteristics generated divergence between rich and poor, but this effect was not strong enough to counteract the adjustment of earnings back to their permanent level.

This chapter also performed some robustness tests to assess the impact of measurement error on the earnings variable, and attrition and non-reporting in the sample. While the effects of measurement error in earnings are usually minor in the case of aggregate measures of mobility (e.g. taking averages across individuals averages out the measurement error) this problem can lead to serious biases when estimating the regressions performed in Chapter 4. The simulations performed indicate that, while measurement error in the earnings variable can create problems in individual mobility analyses, it is unlikely that the mobility results obtained were mainly driven by the measurement error component of earnings.

A more serious problem with the present data, is the large amount of individuals for whom there is no information in the final period. The main reasons for this loss of information are attrition from the panel and non-reporting of earnings. While it is possible that part of the attrition is random, evidence is provided that rich individuals are less likely to report their earnings, complicating the assessment of their mobility. For all these reasons, the mobility results presented in this dissertation do not necessarily apply to the whole urban population in Mexico.

Finally, Chapter 5 tested whether Mexican urban labor markets were segmented in formal and informal sectors. The estimations showed support for the hypothesis that labor markets are segmented. In other words, the answer to the question of "Are formal sector jobs rationed?" is affirmative. Some factors that positively affected the decision of a worker to apply for a formal sector job were his education and his wealth. On the other hand, the number of children in the household affected negatively this decision. It was also found that the education of the worker was always positively associated with the probability of being hired by a formal sector employer.

The sector assignment model generated somewhat extreme predictions about the extent of segmentation in the labor markets. In particular, it predicted that more than 90% of the current informal sector workers were rationed out of the formal sector. While this number seems too high, it was also found that around 30% of those restricted individuals managed to find a formal sector job within a year. That chapter also assessed what were the potential earnings gains that rationed individuals could experience by moving into the formal sector and what was the *actual* earnings mobility experienced by rationed individuals, if they managed to enter the formal sector after one year. The analysis showed that an individual rationed out of the formal sector would on average experience substantial earnings gains if he were to move to the formal sector. The predicted gains are between 7% and 30% of their initial earnings, depending on the sample analyzed. In practice, when comparing these predictions to the gains actually experienced by those restricted individuals who managed to enter the formal sector after one year, it was found that the actual earnings mobility was smaller than predicted. Although the specification of the econometric model needs further tests, Chapter 5 showed that the issue of segmentation in labor markets in urban Mexico is far from being settled, as previous studies had maintained.

The conclusions reached in this dissertation motivate further research that could give a better picture of the functioning of labor markets in Mexico. In particular, it would be interesting to do a more detailed analysis of the short-run dynamic structure of earnings incorporating the information on quarterly earnings. This study could enrich the model of yearly earnings dynamics here presented, and help in better capturing the effects of measurement error on earnings.

One important aspect, omitted in this dissertation, was the study of aggregate *positional* earnings mobility. According to the author's opinion, this is an important topic that has not been pursued satisfactorily in the literature. Most of the positional mobility literature has focused on comparing the mobility of an individual to the whole population. However, in practice there is evidence that individuals pay a lot of attention (if not more) to their relative position within small groups to which they relate, e.g. their communities, their peers, etc. It would be interesting to extend the previous analyses in order incorporate this fact.

BIBLIOGRAPHY

Abowd, J. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, vol. 57(2): 411–445.

Abowd, J. and H. S. Farber (1982), "Job Queues and the Union Status of Workers", *Industrial and Labor Relations Review*, vol. 35(3): 354–367.

Aguayo-Tellez, E. (2005), "Income Divergence between Mexican States in the 1990s: Evidence from Micro-Data", Mimeo.

Antman, F. and D. J. Mckenzie (2005a), "Earnings Mobility and Measurement Error: A Pseudo-Panel Approach", World Bank Policy Research Working Paper # 3745.

Antman, F. and D. J. Mckenzie (2005b), "Poverty Traps and Nonlinear Income Dynamics with Measurement Error and Individual Heterogeneity", Mimeo.

Atkinson, A. B., F. Bourguignon, and C. Morrison (1992), *Empirical Studies of Earnings Mobility*, Chur: Harwood Academic Publishers.

Attanasio, O. P. and M. Székely (2004), "Wage shocks and consumption variability in Mexico during the 1990s", *Journal of Development Economics*, vol. 73: 1–25.

Basch, M. and R. D. Paredes-Molina (1996), "Are there dual labor markets in Chile?: empirical evidence", *Journal of Development Economics*, vol. 50: 297–312.

Basu, K. (1997), *Analytical Development Economics. The Less Developed Economy Revisited*, Cambridge, MA: MIT Press.

Baulch, B. and J. Hoddinott (2000), "Economic Mobility and Poverty Dynamics in Developing Countries", *Journal of Development Studies*, vol. 36: 1–24.

Behrman, J. R., A. Gaviria, and M. Székely (2001), "Intergenerational Mobility in Latin America", Research Department Working Paper # 452.

Benabou, R. and E. Ok (2001), "Mobility as Progressivity: Ranking Income Processes According to Equality of Opportunity", NBER Working Paper # 8431.

Binder, M. and C. Woodruff (2002), "Inequality and Intergenerational Mobility in Schooling: The Case of Mexico", *Economic Development and Cultural Change*, vol. 50: 249–267.

Boeke, J. (1953), *Economics and Economic Policy in Dual Societies*, New York: Institute of Pacific Relations.

Bound, J., C. Brown, G. Duncan, and W. Rodgers (1994), "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data", *Journal of Labor Economics*, vol. 12: 345–368.

Bound, J., C. Brown, and N. Mathiowetz (2001), "Measurement Error in Survey Data", in *Handbook of Econometrics*, (eds.) E. Leamer and J. Heckman, New York: North-Holland, vol. 5 of *Handbooks in Economics*.

Bound, J. and A. B. Krueger (1991), "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?", *Journal of Labor Economics*, vol. 9: 1–24.

Buchinsky, M., G. S. Fields, D. Fougère, and F. Kramarz (2003), "Francs or Ranks? Earnings Mobility in France, 1967-1999", Mimeo.

Cain, G. (1976), "The Challenge of Segmented Labor Market Theories to Orthodox Theory: A Survey", *Journal of Economic Literature*, vol. 14(4): 1215–1257.

Carneiro, P., K. Hansen, and J. J. Heckman (2003), "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice", *International Economic Review*, vol. 44(2): 361422.

Chakravarty, S., B. Dutta, and J. Weymark (1985), "Ethical Indices of Income Mobility", *Social Choice and Welfare*, vol. 2: 1–21.

Coondoo, D. and B. Dutta (1990), "Measurement of Income Mobility: An Application to India", in *Economic Theory and Policy: Essays in Honor of Dipak Banerji*, (eds.) B. Dutta, S. Gangopadhyay, D. Mookherjee, and D. Ray, Bombay and New York: Oxford University Press.

Cortés, F. (2000), *La distribución del ingreso en México en épocas de estabilización y reforma económica*, Mexico: CIESAS & Miguel Ángel Porrua.

Dahan, M. and A. Gaviria (1999), "Intergenerational Mobility in Latin America", Research Department Working Paper # 395.

Deaton, A. (1997), *The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*, Baltimore: The John Hopkins University Press.

Dickens, W. T. and K. Lang (1985), "A Test of Dual Labor Market Theory", *American Economic Review*, vol. 75(4): 729–805.

Doeringer, P. B. and M. J. Piore (1971), *Internal Labor Markets and Manpower Analysis*, Lexington: Lexington Books.

Duncan, G. J. and D. H. Hill (1985), "An Investigation of the Extent and Consequences of Measurement Error in Labor Economics Survey Data", *Journal of Labor Economics*, vol. 3: 508–522.

Duncan, G. J. and N. A. Mathiowetz (1985), *A Validation Study of Economic Survey Data*, Ann Arbor: University of Michigan, Institute for Social Research.

Fields, G. S. (1975), “Rural-Urban Migration, Urban Unemployment and Underemployment, and Job Search Activity in LDC’s”, *Journal of Development Economics*, vol. 2: 165–188.

Fields, G. S. (1990), “Labour Market Modeling and the Urban Informal Sector: Theory and Evidence”, in *The Informal Sector Revisited*, (eds.) D. Turnam, B. Salomé, and A. Schwarz, Paris: Development Centre of the Organisation for Economic Co-Operation and Development.

Fields, G. S. (2001), *Distribution and Development. A New Look at the Developing World*, Cambridge, MA: Russell-Sage Foundation/MIT Press.

Fields, G. S. (2004a), “Does Income Mobility Equalize Longer-Term Income? New Measures of an Old Concept”, Mimeo.

Fields, G. S. (2004b), “Dualism in the Labor Market: A Perspective on the Lewis Model After Half a Century”, *The Manchester School*, vol. 72(6): 724–735.

Fields, G. S. (2004c), “Economic and Social Mobility Really Are Multifaceted”, Mimeo.

Fields, G. S., P. Cichello, S. Freije, M. Menéndez, and D. Newhouse (2003a), “For Richer or for Poorer? Evidence from Indonesia, South Africa, Spain and Venezuela”, *Journal of Economic Inequality*, vol. 1: 67–99.

Fields, G. S., P. Cichello, S. Freije, M. Menéndez, and D. Newhouse (2003b), “Household Income Dynamics: A Four Country Story”, *Journal of Development Studies*, vol. 40: 30–54.

Fields, G. S., J. Leary, and E. Ok (2002), “Stochastic Dominance in Mobility Analysis”, *Economic Letters*, vol. 75: 333–339.

Fields, G. S. and E. Ok (1999a), “The Measurement of Income Mobility: An Introduction to the Literature”, in *Handbook of Income Inequality Measurement*, (ed.) J. Silber, Boston: Kluwer Academic Publishing, vol. 71 of *Recent Economic Thought*.

Fields, G. S. and E. Ok (1999b), “Measuring Movement of Incomes”, *Economica*, vol. 66: 455–472.

Fields, G. S., M. L. Sánchez-Puerta, R. Duval-Hernández, and S. Freije (2005), “Earnings Mobility in Argentina, Mexico, and Venezuela: Testing the Divergence of Earnings and the Symmetry of Mobility Hypotheses”, Mimeo.

Frank, R. H. (1992), "Frames of Reference and the Intertemporal Wage Profile", in *Choice Over Time*, (eds.) G. Loewenstein and J. Elster, New York: Russel Sage Foundation.

Freije, S. (2001), *Household Income Dynamics in Venezuela*, Ph.D. thesis, Cornell University, Department of Economics.

Friedman, M. and S. Kuznets (1954), *Income from Independent Professional Practice*, New York: N.B.E.R.

Furnivall, J. (1939), *Netherlands India: A Study of Plural Economy*, Cambridge: Cambridge University Press.

Gaiha, R. (1988), "Income Mobility in Rural India", *Economic Development and Cultural Change*, vol. 36: 279–302.

Galasi, P. (1998), "Income Inequality and Mobility in Hungary, 1992-96", Innocenti Occasional Papers, Economic and Social Policy Series, no. 64.

Gindling, T. (1991), "Labor Market Segmentation and the Determination of Wages in the Public, Private-Formal, and Informal Sectors in San José, Costa Rica", *Economic Development and Cultural Change*, vol. 39: 585–605.

Glewwe, P. and G. Hall (1998), "Are some groups more vulnerable to macroeconomic shocks than others? Hypothesis tests based on panel data from Peru", *Journal of Development Economics*, vol. 56: 181206.

Gong, X. and A. van Soest (2002), "Wage differentials and mobility in the urban labour market: a panel data analysis for Mexico", *Labour Economics*, vol. 9: 513–529.

Gong, X., A. van Soest, and E. Villagómez (2000), "Mobility in the Urban Labour Market: a Panel Data Analysis for Mexico", Discussion Paper Series no. 213.

Grootaert, C., R. Kanbur, and G.-T. Oh (1997), "The Dynamics of Welfare Gains and Losses: An African Case Study", *Journal of Development Studies*, vol. 33(5): 635–657.

Gunning, J. W., J. Hoddinott, B. Kinsey, and T. Owens (2000), "Revisiting Forever Gained: Income Dynamics in the Resettlement Areas of Zimbabwe, 1983-96", *Journal of Development Studies*, vol. 36: 131–154.

Harris, J. R. and M. P. Todaro (1970), "Migration, Unemployment and Development: A Two-Sector Analysis", *American Economic Review*, vol. 60(1): 126–142.

Hause, J. C. (1977), "The Covariance Structure of Earnings and the On-The-Job-Training-Hypothesis", *Annals of Economic and Social Measurement*, vol. 6(4): 1–24.

Heckman, J. J. and B. Honoré (1990), "The Empirical Content of the Roy Model", *Econometrica*, vol. 58(5): 1121–1149.

Heckman, J. J. and V. J. Hotz (1986), "An Investigation of the Labor Market Earnings of Panmanian Males. Evaluating the Sources of Inequality", *Journal of Human Resources*, vol. XXI(4): 507–542.

Herrera, J. (1999), "Ajuste económico, desigualdad y movilidad", in *Pobreza y economía social. Análisis de una encuesta (ENNIV-1997)*, (eds.) R. Webb and M. Ventocilla, Lima: USAID, Instituto Cuanto, UNICEF.

Ibarlucea, P. (2003), "La movilidad de trabajadores entre sectores económicos: análisis con cadenas de Markov", *Revista Mexicana del Trabajo y la Previsión Social*, vol. 3: 71–137.

Jalan, J. and M. Ravallion (2000), "Is Transient Poverty Different? Evidence for Rural China", *Journal of Development Studies*, vol. 36: 82–99.

Latapí, A. E. (1992), "Men's and Women's Patterns of Intragenerational Occupational Mobility During Mexico's Boom and Crisis", in *The Sociodemographic Effects of the Crisis in Mexico*, (eds.) H. A. Selby and H. Browning, Austin: University of Texas at Austin.

Lewis, A. W. (1954), "Economic Development with Unlimited Supplies of Labour", *Manchester School*, vol. 22: 139–191.

Lillard, L. A. (1999), "Job Turnover Heterogeneity and Person-Job-Specific Time-Series Wages", *Annales d'Économie et de Statistique*, vol. 55-56: 183–210.

Lillard, L. A. and Y. Weiss (1979), "Components of Variation in Panel Earnings Data: American Scientists 1960-70", *Econometrica*, vol. 47(2): 437–454.

Lillard, L. A. and R. J. Willis (1978), "Dynamic Aspects of Earning Mobility", *Econometrica*, vol. 46(5): 985–1012.

López-Acevedo, G. and A. Salinas (2000), "How Mexico's financial crisis affected income distribution", World Bank Policy Research Working Paper # 2406.

Lustig, N. and M. Székely (1999), "Tendencias 'Ocultas' en la Desigualdad y la Pobreza en México", in *Pobreza y Desigualdad en América Latina*, (eds.) M. Cárdenas and N. Lustig, Santafé de Bogotá: TM Editores - FEDESSARROLLO - LACEA - COLCIENCIAS.

MaCurdy, T. E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics*, vol. 18: 83–114.

Maddala, G. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*, New York: Cambridge University Press.

Magnac, T. (1991), "Segmented or Competitive Labour Markets?", *Econometrica*, vol. 59(1): 165–187.

Maloney, W. F. (1998), "The Structure of Labor Markets in Developing Countries. Time Series Evidence on Competing Views.", Policy Research Working Paper no. 1490.

Maloney, W. F. (1999), "Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico", *The World Bank Economic Review*, vol. 13: 275–302.

Maloney, W. F. (2003), "Informal Self-employment; Poverty trap or decent alternative.", in *Pathways out of poverty, private firms and economic mobility in developing countries*, (eds.) G. Fields and G. Pfeffermann, Boston, MA: Kluwer Academic Publishing.

Maloney, W. F. (2004), "Informality Revisited", *World Development*, vol. 32(7): 1159–1178.

Maloney, W. F. and M. Bosch (2005), "Labor Market Dynamics in Developing Countries: Comparative Analysis using Continuous Time Markov Processes", Mimeo.

Maloney, W. F. and W. Cunningham (2000), "Measuring Vulnerability: Who Suffered in the 1995 Mexican Crisis?", Mimeo.

Maloney, W. F., W. V. Cunningham, and M. Bosch (2004), "The Distribution of Income Shocks during Crises: An Application of Quantile Analysis to Mexico, 1992-95", *World Bank Econ Review*, vol. 18(2): 155–174.

Manski, C. F. (1995), *Identification Problems in the Social Sciences*, Cambridge, MA: Harvard University Press.

Manski, C. F. (2003), *Partial Identification of Probability Distributions*, Springer Series in Statistics, New York: Springer Verlag.

Marcouiller, D., V. R. de Castilla, and C. Woodruff (1997), "Formal Measures of the Informal-Sector Wage Gap in Mexico, El Salvador, and Peru", *Economic Development and Cultural Change*, vol. 45: 367–392.

McKenzie, D. J. (2003), "How do Households Cope with Aggregate Shocks? Evidence from the Mexican Peso Crisis", *World Development*, vol. 31: 11791199.

Meghir, C. and L. Pistaferri (2001), "Income Variance Dynamics and Heterogeneity", IFS WP 01/07.

Navarro-Lozano, S. and P. Schrimpf (2004), "The importance of being formal: testing for segmentation in the Mexican labor market", Mimeo.

Nee, V. (1996), "The Emergence of a Market Society: Changing Mechanisms of Stratification in China", *American Journal of Sociology*, vol. 101: 908–949.

Pisani, M. J. and J. A. Pagán (2003), "Sectoral Queuing in a Transitional Economy: The Case of Nicaragua in the 1990s", *Labour*, vol. 17(4): 571–597.

Pischke, J.-S. (1995), "Evidence on the Validity of Cross-Sectional and Longitudinal Labor Market Data", *Journal of Business and Economic Statistics*, vol. 13(3): 305–314.

Poirier, D. J. (1980), "Partial Observability in Bivariate Probit Models", *Journal of Econometrics*, vol. 12: 209–217.

Pradhan, M. and A. van Soest (1995), "Formal and informal sector employment in urban areas of Bolivia", *Labour Economics*, vol. 2: 275–297.

Pratap, S. and E. Quintin (forthcoming), "Are Labor Markets Segmented in Argentina? A Semi Parametric Approach", *European Economic Review*.

Ranis, G. (forthcoming), "Is Dualism Worth Revisiting", in *Poverty, Inequality, and Development*, (eds.) A. de Janvry and R. Kanbur, Kluwer Academic Publishing, vol. 1 of *Economic Studies in Inequality, Social Exclusion and Well-Being*.

Rosenzweig, M. R. (1988), "Labor Market in Low-Income Countries", in *Handbook of Development Economics*, (eds.) H. Chenery and T. Srinivasan, New York: North-Holland, vol. 1 of *Handbooks in Economics*.

Roy, A. (1951), "Some Thoughs on the Distribution of Earnings", *Oxford Economic Papers*, vol. 3(2): 135–146.

Sánchez-Puerta, M. L. (2005), *Earnings Mobility in Urban Argentina 1995-2003*, Ph.D. thesis, Cornell University, Department of Economics.

Selby, H. A. and H. Browning (eds.) (1992), *The Sociodemographic Effects of the Crisis in Mexico*, Austin: University of Texas at Austin.

Shorrocks, A. F. (1978), "Income Inequality and Income Mobility", *Journal of Economic Theory*, vol. 19: 376–393.

Thomas, M. and L. Vallée (1996), "Labour Market Segmentation in Cameroonian Manufacturing", *Journal of Development Studies*, vol. 32(6): 876–898.

Tunalı, İnsan. (1986), "A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Remigration", in *Research in Labor Economics*, (ed.) R. G. Ehrenberg, Greenwich, Conn., and London: JAI Press Inc., vol. 8 (Part B), pp. 235–283.

Wodon, Q. (2001), "Income mobility and risk during the business cycle. Comparing adjustments in labour markets in two Latin-American countries", *Economics of Transition*, vol. 9(2): 449–461.

World Bank (2004), *Poverty in Mexico: An Assessment of Conditions, Trends, and Government Strategy*, Washington, D.C.: The World Bank.

Yitzhaki, S. and Q. Wodon (2002), "Mobility, Inequality, and Horizontal Equity", Mimeo.